SOLID STATE
ELECTRONICS

MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

# REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188



**1. AGENCY USE ONLY** (Leave blank)

**2. REPORT DATE**
March 1, 1996

**3. REPORT TYPE AND DATES COVERED**
Annual 1 March 1995 through 29 February 1996

**4. TITLE AND SUBTITLE**
JSEP Annual Progress Report No. 2

**5. FUNDING NUMBERS**
DAAH04-94-G-0058

**6. AUTHOR(S)**
J. S. Harris, Program Director

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Stanford University
Solid State Electronics Laboratory
CIS-X 329
Stanford, CA 94305-4075

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
ARO 32283.2-EL-JSEP

**11. SUPPLEMENTARY NOTES**
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

**12a. DISTRIBUTION/AVAILABILITY STATEMENT**
Approved for public release; distribution unlimited.

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** (Maximum 200 words)

**14. SUBJECT TERMS**

**15. NUMBER OF PAGES**
69

**16. PRICE CODE**

# JSEP ANNUAL REPORT

1 March, 1995 through 29 February, 1996

James S. Harris, Jr.
JSEP Principal Investigator
and Program Director

(415)723-9775

This work was supported by the
Joint Services Electronics program
(U.S. Army, U.S. Navy and U.S. Air Force)
Contract DAAH04-94-G-0058
and was monitored by the
U.S. Army Research Office

**Abstract**

This is the annual report of the research conducted at the Stanford Electronics Laboratories under the sponsorship of the Joint Services Electronics Program from March 1, 1995 through February 29, 1996. This report summarizes the areas of research, identifies the most significant results and lists the dissertations and publications sponsored by contract DAAH04-94-G-0058.


## Table of Contents


Introduction and Overview of Principal Accomplishments 3

Unit 1: Investigation of Transport in Quantum Dots 7

Unit 2: Patterned Thin-Film Media for High Density Magnetic Recording 15

Unit 3: Investigation of a Metal Source and Drain Field Emission Transistor 20

Unit 4: On-chip Thin Film Solid State Micro-battery 27

Unit 5: CVD Epitaxial Germanium *n*-channel FETs Formed on Si Substrates using Strain-relief Layers 30

Unit 6: Portable Video on Demand in Wireless Communication 39

# JSEP ANNUAL REPORT
## March 1, 1995 - February 29, 1996

### Introduction and Overview of Principal Accomplishments

This annual report covers research accomplishments for the period 1 March, 1995 through 29 February, 1996 for basic electronics research conducted in the JSEP program in the Electrical Engineering Department of Stanford University. The Stanford Electronics Lab JSEP Director and Principal Investigator is Professor James Harris. The program work units are as follows:

| | |
|---|---|
| Unit 1: | Investigation of Transport in Quantum Dots<br>(James S. Harris) |
| Unit 2: | Patterned Thin Film Media for High Density Magnetic Recording<br>(R. Fabian W. Pease) |
| Unit 3: | Investigation of a Metal Source and Drain Field Emission Transistor<br>(C. Robert Helms) |
| Unit 4: | On-chip Thin Film Solid State Micro-battery<br>(S. Simon Wong) |
| Unit 5: | CVD Epitaxial Germanium *n*-channel FETs Formed on Si using Strain-relief Layers<br>(Krishna Saraswat) |
| Unit 6: | Portable Video on Demand in Wireless Communication<br>(Teresa H. Y. Meng) |
| Unit 7: | Adaptive DFE for GMSK in Indoor Radio Channels<br>(John M. Cioffi) |
| Unit 8: | Robust Estimation Methods for Adaptive Filtering<br>(Thomas Kailath) |
| Unit 9: | Efficient Data Compression<br>(Thomas M. Cover) |

### Highlights

In work unit 1, Professor Harris and students have developed the nanofabrication techniques for large (200X200) arrays of 100nm quantum dots and demonstrated the first Coulomb blockade and hysteretic switching behavior in such large arrays. This work represents a significant advance in nanofabrication and demonstrates the robustness of Coulomb blockade compared to quantum interference effects.

In work unit 2, Professor Pease and students have demonstrated and characterized (with Magnetic AFM, alternating gradient magnetometer) magnetic thin film recording media patterned into deep submicron islands for improved density (>12 Gbytes/sq. in.) and lower transition noise. One medium was Polycrystalline Co 20nm thick on Cr which exhibited 1 bit/1 domain/1 island for dimensions less than 150nm. Another medium was single crystal iron film which, when patterned, demonstrated single domain/island behavior for large (1-micron) islands. Magnetic anisotropy in the iron films was dominated by crystalline orientation which allows us to decouple the magnetic direction from the shape of the island; this is valuable for applications involving horizontal recording.

In work unit 5, Professor Saraswat and his students are developing a technology to fabricate high-performance n-channel heterostructure field-effect devices using germanium-rich GeSi grown via graded-alloy strain reduction on (001) silicon substrates. The goal is to combine the high

hopes that more fundamental work ultimately has a greater impact because it leads to things that simply would not have been done if left to only research programs with nearer term, clearly identified needs. The transfers of technology described below are thus the result of JSEP supported programs of 5-10 years ago.

Research into the engineering of silicon nanopillars in Professor Pease's JSEP program has led to new insights into the oxidation of silicon under high stress, confined geometry conditions. As Si ULSI continues to shrink, such high stresses are quite important. The results of this research are now being incorporated into SUPREM process models being developed to simulate the processing of next-generation, ultra-small geometry ULSI circuitry.

An essential element in manufacturing high performance AMLCDs is the ability to fabricate TFT driver circuits and integrate them with the liquid crystals on glass substrates. However, the high temperatures and long thermal cycles generally needed to obtain high performance TFTs cause warpage and shrinkage to glass. As a result, fabrication processes are limited to low temperatures and short times. Early work of Professor Krishna Saraswat funded by JSEP and subsequently by DARPA demonstrated high performance TFTs in poly-GeSi with low thermal budget processing, compatible with glass substrates. He demonstrated significantly lower processing temperatures for deposition, doping, recrystallization, and grain boundary passivation. Several novel device structures have been developed to improve TFT performance, such as, increased drive current in the ``on" state and reduced leakage in the ``off" state. He is actively working with XEROX and Intevac to transfer this technology and several major organizations around the world are now developing the poly-GeSi TFT technology which originated under JSEP support in his laboratory.

The early JSEP work demonstrating the first MBE growth and growth induced layering of the high temperature superconductors by MBE in Professor Harris's program is the basis for the continuing high $T_c$ program at Varian Associates. The focus of their effort is MBE growth induced layering of alternate superconducting and insulating phases to produce well controlled Josephson junctions.

One of the key problems facing modern ultra-high bandwidth communications systems is how to handle the final 100 meters where information delivery is to only a single receiver and the costs of high bandwidth solutions can no longer be divided by a large number of receivers. The early JSEP supported research under Prof. John Cioffi led to the development of the "Discrete MultiTone" (DMT) technology that is now an international standard (ANSI T1.413) for both video transmission and high-speed internet access on twisted pairs, in what is known as Asymmetric Digital Subscriber Lines (ADSL). Stanford holds 4 patents in the area that are exclusively licensed and sublicensed by Stanford to a DMT-spinout, Amati Communications Corporation. Amati has sublicensed the DMT patents to a number of semiconductor and telecom manufacturers around the

world, including Motorola, Northern Telecom, and AT&T (now Lucent Technology). Amati builds products based on the DMT technology and has been extremely successful.

The early JSEP supported work of Professor Tom Cover is now being utilized in many of the WWW browsers. One of the issues is do you wait for all of the information to be supplied serially or do you send information at various levels of refinement so that the description efficiency is optimal at each level ? The idea is to utilize methods of successive refinement to quickly produce a rough picture, then successively more refined pictures until the final version is produced. This

UNIT: 1

TITLE: Investigation of Transport in Quantum Dots

PRINCIPAL INVESTIGATOR: J. S. Harris, Jr.

GRADUATE STUDENTS: D. R. Stewart and C. I. Duruöz

## 1. Scientific Objectives

The continuing drive for increased device density in both IC and memory technologies demands smaller and closer packed future devices. We are pursuing an investigation into the electronic transport in both single quantum devices and large arrays of densely packed quantum dots. A full understanding in both regimes will be required in any successful implementation of single electron electronics. In particular, most studies of quantum devices have concentrated on the very low bias equilibrium behavior [Beenakker][Kouwenhoven]; we concentrate instead on the technologically relevant non-linear high bias operating regime.

We have two main objectives: first, to understand the mechanisms controlling electron transport through single quantum point contacts and quantum dots and second, to study the fundamental characteristics of coulomb blockade and charge coupling in transport through quantum dot arrays.

## 2. Summary of Research

### 2.1 Introduction

We previously reported our initial investigations of the electronic transport through 200 x 200 two dimensional quantum dot arrays patterned on a molecular beam epitaxy (MBE) grown GaAs/AlGaAs heterostructure [Harris][Duruöz]. The current-voltage (I-V) relation of the arrays showed two striking features: a threshold for conduction, and multiple switching events accompanied by a hierarchy of hysteresis loops. By changing the voltage applied to a single Schottky gate deposited over the entire array, it was possible to move between the hysteretic and non-hysteretic regimes. A single hysteresis loop was measured in the single control dots fabricated adjacent to the large arrays. No switching or hysteresis was observed above a temperature of 700mK.

We have continued this investigation by focusing on the mechanisms responsible for the switching and hysteresis. It is this behavior, and control of it, that will be most relevant in any technological application.

We have thus characterized in detail the behavior of the single control quantum dots and point contacts in our first generation devices. We have also fabricated a second generation of similar etch defined single devices using a GaAs/AlGaAs heterostructure grown by chemical vapor deposition. All of our single device results have been duplicated in both of these materials to prove the repeatability and robustness of the switching phenomena. Our results show the single hysteresis observed to be the experimental realization of a basic conduction bistability in the I-V relation. When measured on sufficiently fast time scales, the switching bistability manifests as a random telegraph signal in the current under constant voltage bias. Most significantly, we are able to control the bistable switching rate and range with voltages applied to a new back gate and the original front Schottky gate. We are also able to observe the switching in the new devices at a temperature of 4.2K. These results have yielded new insight into the cause of the I-V switching.

## 2.2 Device Fabrication and Measurement Configuration

All devices measured were fabricated by lithographically patterning a GaAs/AlGaAs epitaxially grown heterostructure. We have utilized a standard modulation doped architecture to create a two dimensional electron gas (2DEG) approximately 800 Å below the wafer surface. First generation and second generation split gate devices were fabricated from MBE material grown in our laboratory with a mobility and sheet density of $\approx 200\,000$ $cm^2/Vs$ and $3.5 \times 10^{11}$ $cm^{-2}$; second generation etched devices were patterned on CVD material grown at Sandia National Labs by our collaborator H.Chui with a mobility and density of $\approx 300\,000$ $cm^2/Vs$ and $2.0 \times 10^{11}$ $cm^{-2}$.

The devices were formed using electron-beam lithography to define the point contact, dot and array features. Minimum feature size as shown in Fig. 1 is 100 nm, point contact barrier openings are 200-400 nm, and the array periodicity is 800 nm. This lithographic pattern was used as a mask for wet chemical etching 800 Å deep through the 2DEG in the case of etched structures, or NiCrAu metal gate evaporation for the split gate devices. A single 1000 Å Au front gate was deposited over the etched devices. A ground plane below the mounted chips was used as a back gate.


Gate
Control

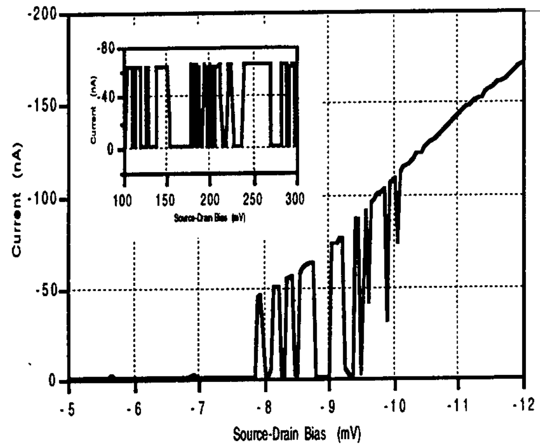



Figure 3: I-V curve of quantum dot displaying bistable conduction switching as the bias is swept up over 8-10 mV. Inset shows random telegraph signal in time at a fixed bias of -9 mV. Temperature is 400 mK.

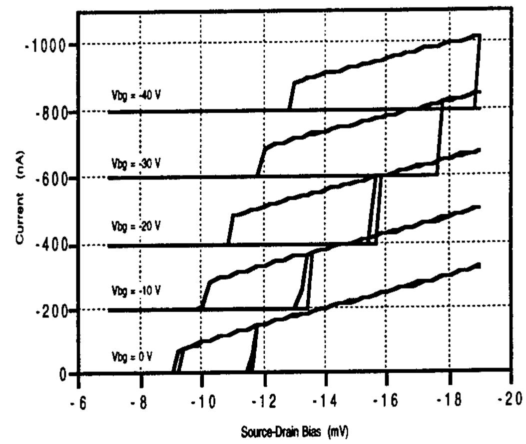



Figure 4: In the hysteretic regime, control over the size and position of the hysteresis loop is effected with a backgate voltage as labeled (curves offset for clarity). Results for an etch defined quantum dot at 400 mK.

As the source-drain bias is swept over the switching range these lifetimes appear to change exponentially; $t_{high}$ increases with bias and $t_{low}$ decreases. The clean hysteresis loops initially observed in the arrays can thus be described as bistable conductance regions with average ( $t_{high}$, $t_{low}$ ) » measurement sweep rate. As the device remains cold, the time constants of this switching increase over several hours until even a slow voltage sweep appears hysteretic.

In this long switching time or 'hysteretic' regime when $t_{switch}$ is much greater than our measurement speed of O( 10s ), we can use the front and back gates to control the size and position of the hysteresis. As an increasingly negative backgate voltage is applied, the hysteresis loop expands in size and the initial turn on threshold shifts to higher source-drain bias, as illustrated in Fig. 4.

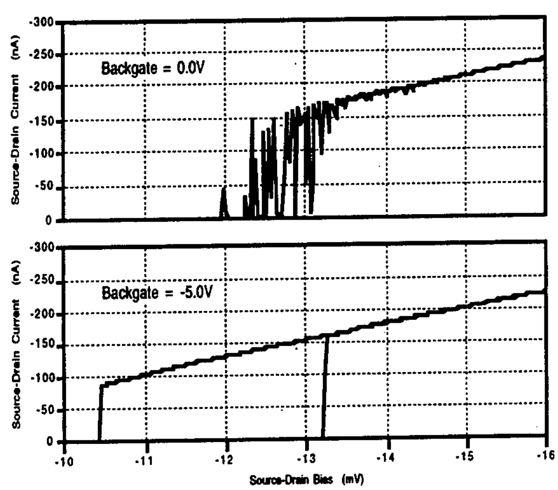



Figure 5: In the fast telegraph switching regime, the backgate is able to reversibly control the bistable state lifetimes. Results from etch defined quantum dot at 400 mK.

In the short switching time or telegraph noise regime we achieve our most significant result; application of a small backgate voltage changes the average state lifetimes dramatically. We are able to continuously control the lifetimes over our full measurement range of 100μs to 1000s, seven orders of magnitude. Fig. 5 demonstrates this control.

The CVD etched devices extended the temperature range of this behavior to above 4.2K. In addition, some of these devices displayed multi-stable switching instead of a simple bistability. The multi-stable devices also showed switching between finite conduction states, and a smoother current turn on. This comparison is made in Fig. 6.

We have also conducted initial tests on the split gate second generation single devices, in

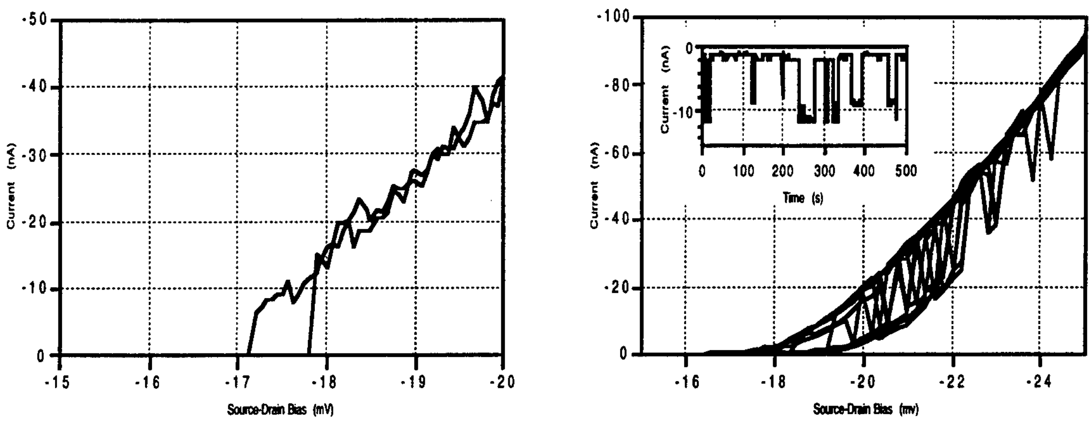



Figure 6: (a) Bistable hysteresis in a CVD point contact at 4.2 K (b) Multi-stable switching and associated multi-level random telegraph signal (inset) in another CVD point contact at 4.2 K.

which the quantum barriers are defined with electrostatic depletion gates instead of wet chemical etching. Well resolved coulomb blockade measurements (Fig. 7) demonstrate that these devices are performing correctly. Future measurements will characterize and compare the switching behavior in this very different architecture to the etched device results.

## 2.4 Discussion of the Results

The most significant result in the single device investigation has been the characterization of the hysteresis as a basic conduction bistability with a random telegraph signal (RTS). This result has been confirmed in the high bias regime by other groups in an offset split gate [Smith] and a deeply etched lateral barrier [Pilling]. Random telegraph signals have been observed in quantum devices near equilibrium [Dekker][Timp][Sakamoto] and have been attributed to the fluctuations of a single or small number of nearby impurities. Many of our results are consistent with this interpretation, however the exponential dependence of the bistable state lifetimes $t_{high}$, $t_{low}$ as a function of source drain bias has not been measured before, and remains difficult to interpret

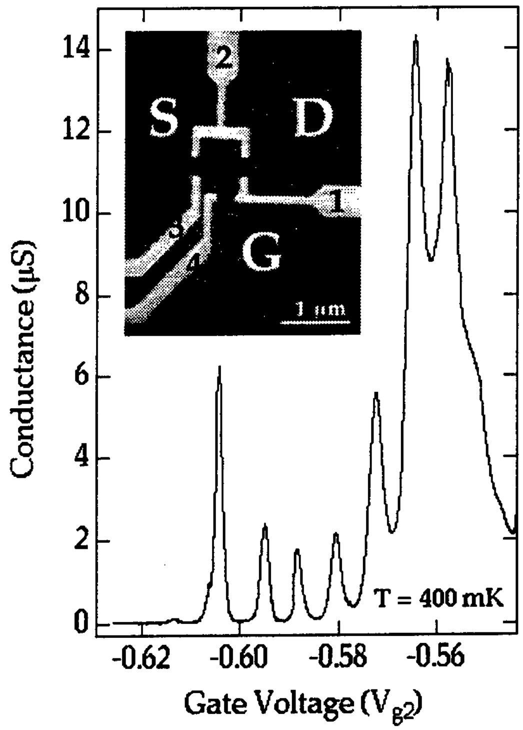



Figure 7: Coulomb blockade oscillations in a three lead dot. The inset shows the SEM picture of the device. Top gates are numbered from "1" to "4". "G", "D" and "S" denote the semi-infinite leads can be used interchangeably as "Source", "Drain" and "Leakage Channel". The result shown here is obtained by varying the voltage on gate "2", and keeping the others constant.

within the impurity model. The very strong control effected by the back gate voltage on switching times is likewise unexplained.

The multi-stability displayed in some of the CVD etched devices (Figure 6) is more typical of fluctuations due to impurities. Yet in this case as before there is an exponential bias dependence of lifetimes, and indeed under controlled circumstances an evolution from bistable on-off switching to multi-stable on-on transitions.

Voltage dependent random telegraph signals have been observed in submicron MOSFET inversion layers and 4μm diameter resonant tunnel diodes [Ralls][Ng]. In each case the dependence of the RTS is attributed to the physical position of a switching impurity and it's bias defined energy with respect to a local Fermi level. In our devices the voltage dependence scale is much smaller - the state lifetimes can vary by two orders over only 500 μV of applied bias, inconsistent with the above explanation.

## 3. Conclusions and Future Work

The cause of the conduction instability remains unclear. Strong qualitative similarities to impurity switching results are contradicted by the exponential voltage dependencies of the state lifetimes. However, we have already been able to demonstrate remarkable control over the character of the instability as it manifests in the I-V relation using both front and back gate potentials. Further probing of this control should lead to a physical explanation of the switching and hysteresis.

We will continue with a series of measurements characterizing the transition from the well understood equilibrium regime to our high bias non-equilibrium situation. Quantitative dependencies of the state lifetimes as a function of gate voltages, applied bias and temperature across this transition are required. Similar measurements on our split gate devices will quantify the relevance of the surfaces and associated imperfections in the etched devices, and direct future fabrication towards the most robust architecture.

With these results in hand, we will return to the performance of the single device arrays, densely packing the point contacts and quantum dots into 1D and 2D arrays. Single and coupled device behavior can then be separated and accurately characterized. This knowledge will form the design framework of future single electron architectures in this regime.

## 4. References


[Beenakker] C. W. J.Beenakker *et al., Phys. Rev. B* 44, 1646 (1991)

[Dekker] C. Dekker *et al., Phys. Rev. Lett.* 66, 2148 (1991)

[Duruöz] C. I. Duruöz *et al., Phys. Rev. Lett.* 74, 3237 (1995)

[Harris] J. S. Harris Jr. *et al., JSEP Annual Report* (1994-1995)

[Kouwenhoven] L. P. Kouwenhoven *et al., J. Phys. B - Cond. Matt.* 85, 367 (1991)

[Ng] S.-H. Ng *et al., Appl. Phys. Lett.* 62, 2262 (1993)

[Pilling] G. Pilling *et al., Proceedings EP2DS XI* 347 (1995)

[Ralls] K. S. Ralls *et al., Phys. Rev. Lett.* 52, 228 (1984)

[Sakamoto] T. Sakamoto *et al., Appl. Phys. Lett.* 67, 2220 (1995)

[Smith] J. C. Smith *et al., Proceedings EP2DS XI* 351 (1995)

[Timp] G. Timp *et al., Phys. Rev. B* 42, 9259 (1990)


## 5. JSEP Supported Publications


1. C. I. Duruöz, R. M. Clarke, C. M. Marcus and J. S. Harris Jr., "Conductance Threshold, Switching and Hysteresis in Quantum Dot Arrays," *Phys. Rev. Lett.* 74, 3237 (1995).

2. C. I. Duruöz, D. R. Stewart, C. M. Marcus and J. S. Harris Jr., "Switching and Hysteresis in Quantum Dot Arrays," *Proceedings EP2DS XI* 349 (1995).
3. G. Pilling, D. H. Cobden, P. L. McEuen, C. I. Duruöz and J. S. Harris Jr., "Intrinsic Bistability in Nonlinear Transport Through a Submicron Lateral Barrier," *Proceedings EP2DS XI* 347 (1995).
4. G. S. Solomon, C. I. Duröz, C.M. Marcus and J. S. Harris, Jr., "Growth Induced and Patterned 0-Dimensional Quantum Dot Structures" in Low Dimensional Structures Prepared by Epitaxial Growth or Regrowth on Patterned Substrates, ed. by K. Eberl et al., NATO ASI Series E, Applied Sciences 298.

## 6. JSEP Supported Ph. D. Thesis

C. I. Duröz, "Low Temperature Transport in Quantum Dot Arrays", Ph. D. Thesis, Stanford University, March, 1996.

UNIT: 2

TITLE: Patterned Thin Film Media for High Density Magnetic Recording

SENIOR INVESTIGATOR: R. F. W. Pease

RESEARCH STUDENT: R. M. H. New

## Background

In conventional hard-disk magnetic recording systems, the signal to noise ratio is often limited by "transition" noise which occurs due to the irregular zig-zag domain walls between adjacent recorded bits [Tong]. In order to address this problem, we are studying recording media composed of large arrays of submicron lithographically defined single-domain magnetic islands. It is known both from theoretical arguments and from experiments that sufficiently small magnetic particles are uniformly magnetized and contain no domain walls. If a single-domain particle of this type has a single uniaxial easy axis of magnetization then it will have only two possible magnetization states and will be ideal for storage of a single bit of information. A magnetic recording medium consisting of an array of equally spaced and uniformly shaped single-domain islands with predictably oriented easy axes could serve as a virtually noise-free alternative to the unpatterned magnetic thin films used in conventional hard disk systems. The ultimate theoretical storage density for such a system would be limited only by the spontaneous thermal switching of bits, a problem that would occur only for particles one hundred angstroms in diameter or less.

In a previous contract period we developed a procedure for patterning polycrystalline magnetic thin films using direct-write electron beam lithography and a multi-step masking and milling process [New (a)]. We used this procedure to define large arrays of 0.15μm by 0.2 μm cobalt islands and studied the physical properties of these islands using atomic force, scanning electron and transmission electron microscopy. The magnetic properties were examined with both magnetic force microscopy and bulk hysteresis loop measurement techniques [New (b)].

For those initial experiments we patterned magnetic islands out of a 200-Å-thick polycrystalline cobalt film. Our results indicated that the transition from the multidomain to single domain state occurs at an island diameter of roughly 0.2mm. The magnetic force microscopy images of these islands showedthat these islands were not single domain. However, smaller islands, roughly 0.15mm by 0.2mm in size, were almost all single domain. Transmission

electron microscopy images of the patterned polycrystalline islands indicated that there were roughly 200 cobalt grains per island, each of which has an easy axis of magnetization randomly oriented in the plane of the film. For islands with only a few hundred grains or less, the magnetocrystalline anisotropies of the individual grains may not completely average out and the net magnetocrystalline anisotropy may be larger than the shape anisotropy for some island geometries. Our calculations indicated that for the island geometries we are using, there is a significant probability that the net easy axis may be misaligned with the long axis of the island [New (c)], and our initial experiments confirmed this. Such unpredictably oriented easy axes would cause problems in a single-bit-per-island recording scheme.

One problem with polycrystalline magnetic recording films, either patterned or unpatterned, is that the fundamental unit of magnetization (typically a single grain or grain cluster of 100 to 500 Å in diameter) is not much smaller than the size of a single recorded bit. For a state of the art 1Gbit/in$^2$ recording system, there may be only a hundred grain clusters or less per bit. Because the medium is so coarsely discretized, conventional magnetic recording systems suffer from increasing

During the reporting period the student, Richard M. H. New, completed his PhD. requirements and graduated and is now at the IBM Almaden Research Center San Jose CA. His dissertation, "Patterned Media for High Density Recording", was approved in September 1995 and copies are available.

**References**


[New (a)] R. M. H. New, R. F. W. Pease, R. L. White, *J. Vac. Sci. Technol. B*, **6**, 3196, Nov/Dec 1994.

[New (b)] R. M. H. New, R. F. W. Pease, R. L. White, *J. Vac. Sci. Technol. A*, May/June 1995.

[New (c)] R. M. H. New, R. F. W. Pease, R. L. White, submitted to IEEE International Magnetics Conference, April 1995.

[Tong] H. C. Tong, R. Ferrier, P. Chang, J. Tzeng and K. L. Parker, *IEEE Trans. Mag.*, **20**, 5, 1831 (1984).


**JSEP Supported Publications**

1. "Magnetic force microscopy of single-domain single-crystal iron particles with uniaxial surface anisotropy," R. M. H. New, R. F. W. Pease, R. L. White, R. M. Osgood, K. Babcock, to be published in the *Proceedings of the 40th Annual Conference on Megnetism and Magnetic Materials (J. Appl. Phys.)* held in Philadelphia, Nov. 1995.
2. "Lithographically patterned single domain cobalt islands for high density magnetic recording," R. M. H. New, R. F. W. Pease, R. L. White, to be published in the *Proceedings of the 6th International Conference on Magnetic Recording Media (J. Magn. Mag. Mater.),* held in Oxford, England, July 1995.
3. "Effect of magnetocrystalline anisotropy in single-domain polycrystalline cobalt islands," *IEEE Trans. Mag.*, MAG-31, p. 3805, Nov. 1995.

**JSEP Support Thesis**

"Patterned Media for High Density Magnetic Recording," R. M. H. New, Ph.D. Thesis, Stanford University, September, 1995.

UNIT: 3

TITLE: **Investigation of a Metal Source and Drain Field Emission Transistor**

**PRINCIPAL INVESTIGATOR: C. R. Helms**

**GRADUATE STUDENT: J. P. Snyder**

## Background

Metal source and drain Metal-Oxide-Semiconductor-Field-Effect-Transistors (MOSFETs) have been shown to have several key advantages over their conventional (doped source and drain) counterparts including ease of fabrication and unconditional immunity to parasitic bipolar and latch-up effects. They were first investigated in the late 1960s [Lepselter], and were thought to have certain advantages over their conventional (diffused source and drain) counterparts including a simplified process, the ability to make very shallow source and drain regions, low source and drain sheet resistance, and complete immunity to latch-up and parasitic bipolar effects. They proved to be poor performers however when compared to a similarly sized conventional MOSFET. The lower drive current in the 'on' state was attributed to the presence of a finite 'gap' between the edge of the poly gate and the edge of the platinum silicide (PtSi) source metal. The much higher leakage currents in the 'off' state originate at the drain end of the device, where electric fields cause the thermally assisted field emission of electrons from the drain into the silicon [Lepselter] [Oh] [Koeneke] [Sugino] [Tsui].

Until recently, the low temperature characteristics of these devices have not been investigated. The only exception to this is a 1968 paper [Lepselter] in which 77 K I-V curves are shown and briefly discussed. Their device was fabricated with a non-self aligned, chemical vapor deposition (CVD) gate oxide process. The data shows a significant *decrease* in current drive at 77 K compared to room temperature

Since 1993, several papers [Tucker] [Hareland] have reported on simulations on these and similar devices, and have shown acceptable drive current and short channel effects in devices with

light of these recent studies to build a metal source and drain device that has all the advantages previously mentioned, as well as superior scalability to well below 0.1 μm and free of the low drive and high leakage current problems. The only requirement is low temperature operation.

## Progress during the current Year

We report the first detailed experimental investigation of the low temperature, field [illegible]

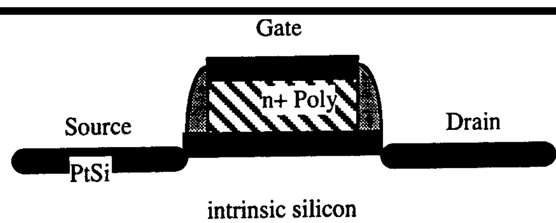



Figure 1: Schematic Diagram of the Device.

various temperatures down to 4.2 K and for channel lengths down to 1 μm. Device fabrication has been optimized so that it is free from the 'gap' at the poly edge described earlier. As will be discussed, we observe a definite transition in the current flow mechanism of the device, from thermal to field emission, as the temperature is reduced below 100 K. In this low temperature 'field emission mode', the drive current when the device is 'on' is comparable to that of a conventional MOSFET, and short channel effects are not observable down to 1 μm, despite the fact that the

channel, as is seen in the 'thermal emission characteristic' drawn in the plot of source current ($I_S$)

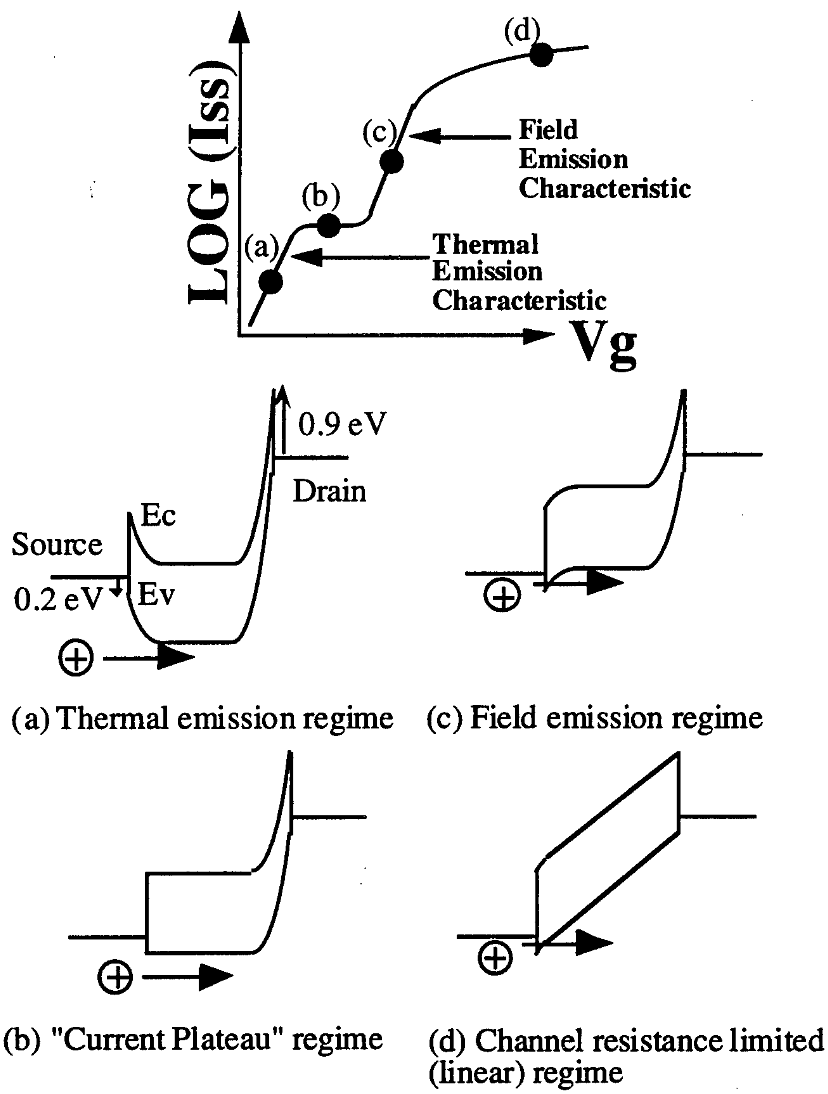



Figure 2. A band diagram description of the different current flow regimes seen in a typical source current vs. gate voltage plot. (a) Thermal emission regime (b) "current plateau" regime (c) field emission regime and (d) channel resistance limited regime.

vs. $V_g$. There is also the possibility of electrons being field emitted from the drain because of the high electric fields there, but this component of current does not show up in our measurements of source current and will not be discussed in this report.

Eventually, with increasingly negative gate bias, only the fixed Schottky part of the barrier to holes remains and the current is limited by thermal emission over this barrier [Fig. 2(b)]. In this

'current plateau' regime further increases in the magnitude of the gate voltage cease to have an exponential effect on Is. The hole current is, for the most part, dependent only on the temperature and the barrier height (~ 0.2 eV), as is drawn in the topmost plot.

With high enough gate bias, holes eventually can be made to tunnel through the Schottky barrier and Is once again begins to increase in an exponential fashion, this time along a 'field emission characteristic' [Fig. 2(c)]. The current is not yet large enough to give the silicon bands in the channel appreciable slope, which is to say that the current is still field emission limited and still travels by diffusion from source to drain, and is not yet channel resistance limited.

Finally $I_s$ becomes large enough that the channel resistance begins to dominate and the holes travel by drift [Fig. 2(d)]. In this regime of $V_g$ the current drive of the device is similar to that of a conventional MOSFET as the Schottky barrier has been rendered all but transparent to the flow of holes.

Drain curves ($I_s$ vs. drain voltage (Vd)) and gate curves ($I_s$ vs. $V_g$) were measured with a computer controlled HP 4140B DC voltage source/pA meter. A Lakeshore cryogenic probe station was used to perform measurements down to 4.2 K.

Figure 3(b) shows the experimental gate curves of the device described in Figs. 1 and 2 with width=length=2 μm. Here the thermal emission, plateau, field emission and channel resistance limited regimes are clearly seen, especially for the 200 K curve. As was mentioned previously, the plateau current is solely a function of temperature and barrier height and this dependence is observable. The plateau current drops exponentially with temperature, so that for temperatures less than about 100 K all significant current flow (> 0.1 pA) occurs by the process

temperatures. This formula gives a barrier of ~0.195 eV, in very good agreement with published

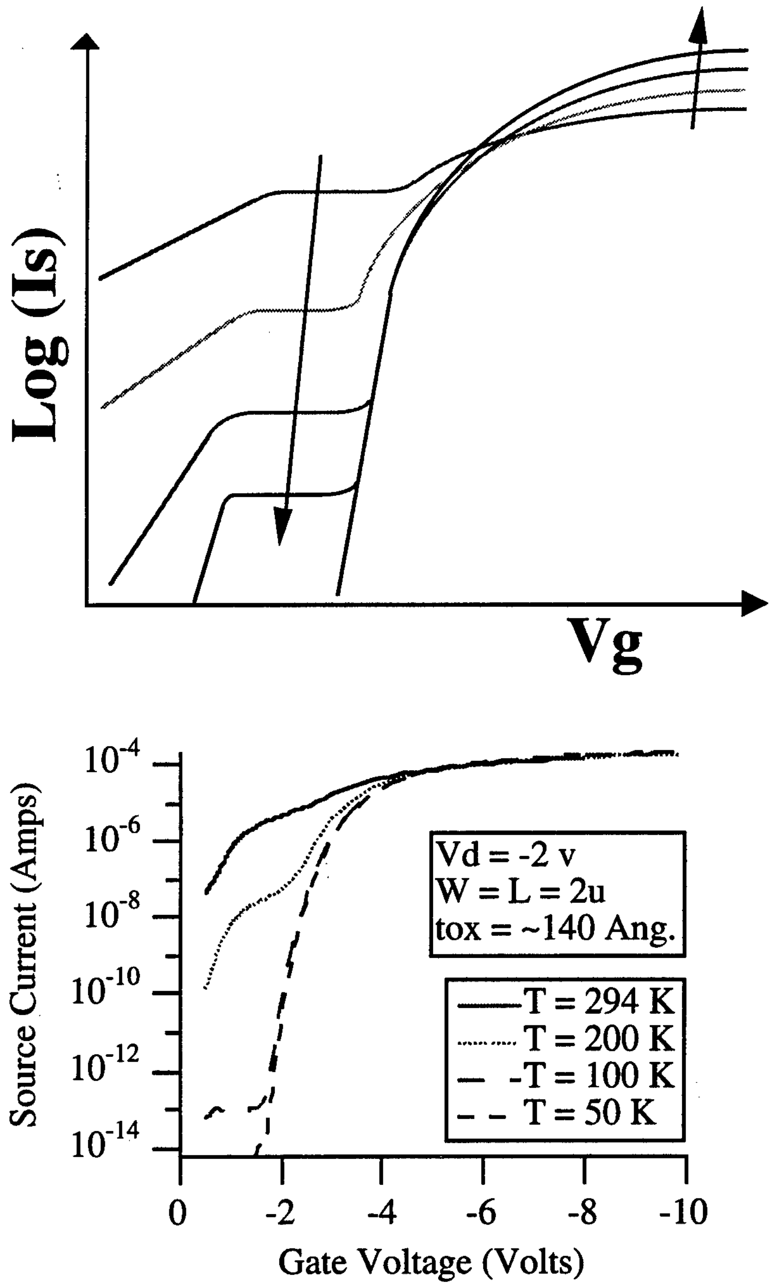



Figure 3. Variation with temperature (a) qualitative example showing the major effects of temperature variation on the gate curves of a PtSi source and drain MOSFET. The arrows point in the direction of decreasing temperature. (b) Actual measured data of a device described in (a).

barrier heights of the PtSi - Si system [Mooney] [Weeks].

During the last year also developed a full 2-D Poisson solver which is integrated with first principles tunneling calculations in order to theoretically examine the effects of device geometry (tip sharpness, channel length, and gate oxide thickness) and materials and system parameters (Schottky barrier height and temperature) on the hole and electron field emission characteristics. The subthreshold slopes of these characteristics were found to decrease monotonically with gate oxide thickness with no theoretical limit. This is in contrast to the theoretical limit, defined by temperature, that exists for the subthreshold region of a conventional device. Subthreshold current levels were also found to be generally smaller than those of conventional devices by several orders of magnitude. Shallow source/drain junctions with sharp tips were found to be optimal in terms of promoting hole field emission drive currents and controlling Drain-Induced-Barrier-Thinning (DIBT) hole leakage currents. Low barrier heights (for good drive currents) and low temperatures (for low leakage over the low barrier) were also found to be optimal.

## Possible Future Directions

These devices will be investigated further. Shallower junction, p+ poly, no gap devices (unlike the ones studied in this dissertation) will be investigated especially with regard to drive current and electron leakage current. NMOS devices can be built as long as metal-silicon Schottky diodes with low barriers to electrons can be found. Rare-earth silicides are potential candidates for this application. Finally, full 2-D modeling of these field emission devices with integrated tunneling and hot-carrier models will be used to further explore the 'virtual source voltage' phenomena described in Chapter 8 of J. P. Snyder's Ph.D. thesis, and to determine the effects of this phenomena on device long term reliability.

## References


[Hareland] S. A. Hareland, A. F. Tasch, C. M. Maziar, *Electronics Letters*, **29**, 1894 (1993).

[Hareland] S. A. Hareland, A. F. Tasch, C. M. Maziar, *Proceedings of the 21st International Symposium on Compound Semiconductors*, September 18-22, San Diego, CA (1994).

[Koeneke] C. J. Koeneke, S. M. Sze, R. M. Levin, E. Kinsbron, 1981 IEDM, 367.

[Lepselter] M. P. Lepselter, S. M. Sze, *Proceedings of the IEEE*, 1400 (1968).

[Mooney] J. M. Mooney, J. Silverman, M. M. Weeks, *SPIE, Infrared Sensors and Sensor Fusion*, **782**, 99 (1987).

[Oh] C. S. Oh, Y. H. Koh, C. K. Kim, 1984 IEDM, 609.

[Sugino] M. Sugino, L.A. Akers, M.E. Rebeschini, 1982 IEDM, 462.

[Tsui] B. Tsui, M. Chen, *J. Electrochem. Soc.*, **136**, 1456 (1989).

[Tucker] J. R. Tucker, C. Wang, J. W. Lyding, T. C. Shen, G. C. Abeln, *1994 SSDM*, 322.

[Tucker] J.R. Tucker, C. Wang, P.S. Carney, *Appl. Phys. Lett.*, **65**, 618 (1994).

[Weeks] M. M. Weeks, P. W. Pellegrini, SPIE, Test and Evaluation of Infrared Detectors and Arrays, **1108**, 31 (1989).


## JSEP Supported Publications

J. P. Snyder and C. R. Helms, Y. Nishi, "Experimental investigation of a PtSi source and drain field emission transistor," *App.Phys.Lett.* **67**(10), 4 September 1995.

UNIT: 4

[redacted]

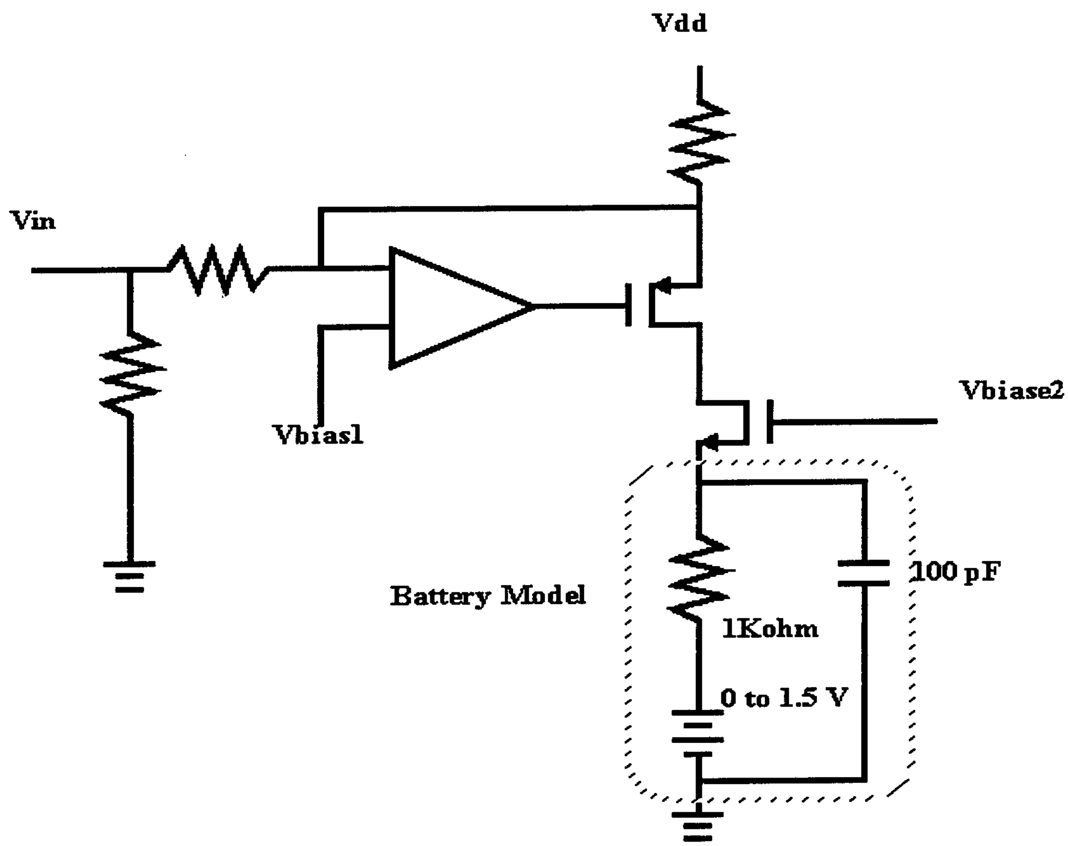



Figure 1. Basic charging circuit.

Figure 2 illustrates the sequence for monolithic integration. The circuits will be first fabricated with a conventional CMOS technology. Afterwards, a layer of silicon oxynitride passivation layer will be deposited using plasma enhanced chemical vapor deposition (PECVD). Lastly, the various layers for the lithium battery will be sputtered on.

The circuits will be fabricate on 4-inch wafers in a 2 μm CMOS technology. Individual die size is limited to about 8 μm by 8 μm. Ten micro-batteries will be sputtered on each wafer. Each battery will be about 1 cm by 1 cm and with a charge capacity of about 1 Coulomb. An overview of the wafer is shown in Fig. 3.

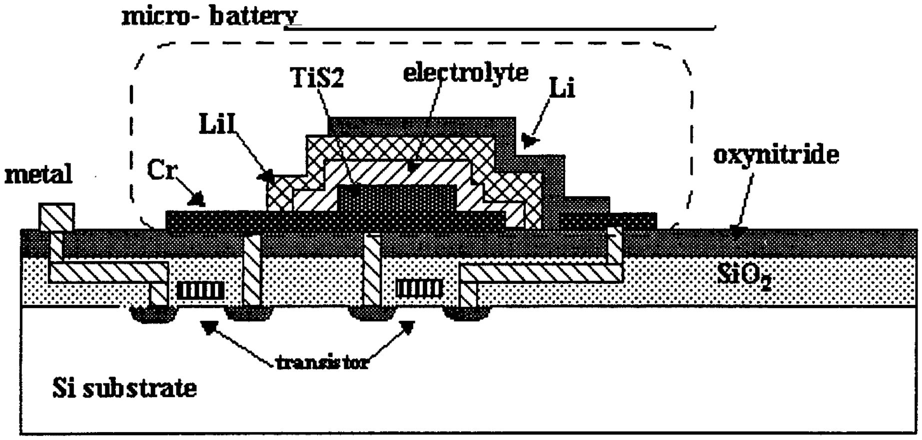



Figure 2. A cross section of the integrated micro-battery.

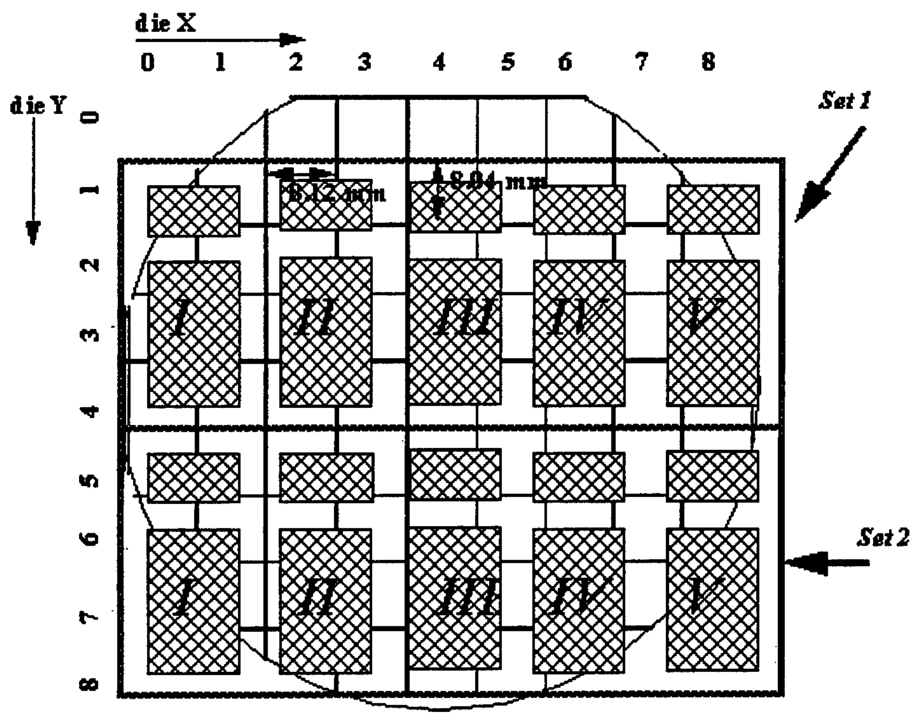



Figure 3. Placement of micro-batteries on a four-inch wafer.

The fabrication will commence shortly. We aim at not only demonstrating the feasibility of such a full integration, but also performing on-chip characterization.

## References

[Mannasiwitsch] V. Manassewitsch, Frequency Synthesizer, Wiley, 3rd Edition.

UNIT: 5

# TITLE: CVD Epitaxial Germanium *n*-Channel FETs Formed on Si Substrates using Strain-relief Layers


PRINCIPAL INVESTIGATOR: K. Saraswat

GRADUATE STUDENT: D. Connelly



## Abstract

*N*-channel field effect transistors are fabricated in strained and unstrained Ge grown via graded-alloy strain reduction on (001) silicon substrates. Applications of Ge device integration with silicon substrates are discussed. Blanket graded-alloy epitaxy is compared with other strain reduction techniques. The effect of strain on the Ge conduction band structure and hence on electron transport in the *x*-*y* plane is examined.


## Objectives

The following are the primary objectives of this project::

- To fabricate *n*-type Ge-channel MOSFETs on a Si substrate.
- To investigate the effect of different degrees of compressive strain on the electron transport properties in germanium inversion layers.
- To compare different schemes for the formation of strain-relief structure formation including blanket graded epitaxy, selective graded epitaxy, and graded epitaxy on ultra-thin silicon-on-insulator.
- To assess the utility of high-germanium content *n*-channel MODFETs in high-speed transistor applications.

## Prior Art

The development of strained layer epitaxy of GeSi alloys on silicon substrates sparked interest in the development of heterostructure devices using silicon-based technology. Much of the work can be placed in one of two categories, vertical heterostructure bipolar transistors (see for example [King]), in which the primary interest is the band-gap difference between the base alloy and the emitter alloy, and confined-carrier field-effect devices (see for example [Pearsall86] and [Daembkes]) in which the parameter of interest is the conduction band offset (for *n*-channel devices) or the valence band offset (for *p*-channel devices).

The biaxial compressive strain formed when GeSi with non-zero $x$ is deposited on silicon enhances the natural positive valence band offset of the GeSi relative to silicon [Pearsall89]

The key difficulty in the formation of these structures is the preparation of the initial [redacted]

should be used. Unfortunately wafers of arbitrary germanium content are not available --- silicon wafers are widely available and germanium wafers are available at considerably higher cost. A solution is to deposit a relaxed "buffer layer" in which threading dislocations are isolated below the surface to translate the surface composition to the desired value from what is on hand. Leaders in this technique include AT&T with Molecular Beam Epitaxy and IBM with CVD and MBE. All reference cases described here begin with (001) silicon wafers.

conduction states in the material. At higher Ge concentrations, however, the strong alloy-dependence of the eight-fold degenerate <111> *L*-valleys brings them to a lower energy.

Due to the dependence of the valence band energy on alloy content across the material spectrum most unipolar heterostructure devices built in the low-Ge regime have used holes as their carrier. *n*-type devices have been built, however, exploiting the strain-dependence of the conduction band minimum.

When (001) silicon is deposited pseudomorphically on a thick unstrained crystalline GeSi alloy the silicon is in biaxial tension, with decreased lattice spacing in the growth direction ($z$) and increased lattice spacing in the two orthogonal directions ($x$ and $y$). The result is that electrons in the $z$-valleys ( [001] and [00-1] ) exhibit a reduced energy relative to those in unstrained silicon while the $x$ and $y$ valleys see an increase in the energy of their states. (See [Pearsall89] for a good overview of the strain effects on GeSi bands.) The advantages are two-fold. First, since the unstrained GeSi substrate has similar conduction band energies to unstrained silicon, the Si now has a reduced conduction band energy relative to the surrounding material and electron confinement can be achieved. The second advantage is that these valleys exhibit a transverse effective mass lower than their longitudinal effective mass. Since conduction in the channel by $z$-valley electrons will be characterized by the lower transverse effective mass while electrons in the other four valleys will be subject to a mixture of the longitudinal and transverse effective masses, preferential occupation of the $z$ valleys results in a decrease in net effective mass and a corresponding increase in mobility for appropriate carrier densities. The stress-induced electron confinement for devices in principle works for alloys from zero Ge up to approximately 80 atomic percent Ge. However, work to date has focused on using strained silicon as the channel material.

In Ge-rich material there is therefore available two mechanisms to yield band offsets. If the unstrained starting material is (001) $Ge_{0.75}Si_{0.25}$ then application of a strained layer of pure Ge will result in a reduced conduction band energy due to the lower energy of the *L*-valleys (due to symmetry the effect of the [001] compression on the <111> *L*-valleys is small). Growth of a strained $Ge_{0.50}Si_{0.50}$ film on the same substrate will result in reduction of the $z$-valley energies relative to the unstrained material. These offsets could be used in the formation of confined-electron structures.

Of further interest in Ge channel devices is in which valleys the conduction band minimum occurs. As the degree of [001] compression is increased via a lowering of the effective substrate germanium content, the energy reduction of the $x$ and $y$ valleys increases the population of

electrons occupying them until they become the principle repository for channel electrons. The effect of this transition on electron mass and electron scattering is of significant importance.

Of practical interest is the formation of the relaxed buffer layer. Linear grades can be done via different temperature schedules to confine stress-relieving defects below the surface. These grades can be executed either on a blanket wafer or in regions defined in a surface oxide layer. Another option is the formation of a graded buffer layer on ultra-thin silicon-on-insulator, decreasing the energy needed to relax the surface.

**Current Work**

While there are many interesting possibilities with Ge-on-Si devices, due to the considerable challenges encountered in the optimization of the graded epitaxial process and in the reliable formation of dielectrics on a germanium surface, this project is focusing on two, both currently under fabrication. One is simple Ge-on-Si *n*-channel field effect transistors. These are expected to exhibit conduction-band minima in the *L*-valleys such as those exhibited by bulk germanium, as was discussed in the last section. The second type of device is the strained Ge-channel on strain-reduced GeSi using a germanium atomic fraction of 75%. It is expected that the strain will reduce the energy of *x* and *y*-directed delta-points below the *L*-valleys, yielding a significant and observable difference in in-plane carrier transport.

Strain-relief via graded epitaxy is achieved by grading the composition, pressure, and temperature in the epitaxial reactor. Depositions are done in the Stanford Center for Integrated Systems Applied Semiconductor Materials Epsilon Chemical Vapor Deposition Epitaxial Reactor. The reactor is a multi-lamp-heated single-wafer unit with a graphite susceptor.

Starting wafers are 4-inch 10 ohm-cm boron-doped (001) silicon. These are cleaned via

98%, however the "discontinuous" jumps from 0 to 3% and 98% to 100% are accommodated without noticeable quality degradation in the film quality.

The key to successful strain relaxation is to maximize the strain reduction achieved via the formation of buried misfit dislocations. These nucleate either homogeneously (thermally) or heterogeneously (due to external factors, such as particles, the wafer edge, etc.). These misfits generally form and propagate in either the [110] or [1-10] direction until either the temperature drops below a kinetic threshold, the edge of the epitaxial region (the wafer edge in the case of blanket epitaxy) is reached, or they scatter towards a wafer surface in the form of threading arms. Since threading arms at the surface can degrade device performance, the distance the misfits are able to travel before scattering should be maximized. A combination of high deposition temperature to drive the propagation kinetics, low deposition rate to give the misfit time to propagate, and low growth rate to maintain an acceptable level of residual strain is thus desirable. Low deposition rate is accomplished by keeping the partial pressures of silane and germane low. However, the combination of a low deposition rate and a gentle alloy gradient yields long deposition times, a potential practical impediment. High deposition temperature causes other problems. Gas phase nucleation, which causes particulate contamination of the surface and formation of a non-epitaxial film, is activated with temperature. Another practical problem with high deposition temperatures is coating of the chamber wall can occur. Since stopping the deposition in-progress is undesirable, it is important that chamber deposition be kept sufficiently low that quartz transparency is maintained.

The primary tools used for material quality determination, other than device fabrication, have been AFM, TEM, RAMAN spectroscopy, EMP, RBS, and anisotropic etches. AFM is of particular interest, as it can be done nondestructively with rapid turnaround on the full-wafer Park Scientific atomic force microscope in the Stanford Center for Integrated Systems. The strain reduction process results in surface undulations in the material. When grading is done from silicon to pure germanium, the peak slope of these undulations is approximately one degree with a mean spacing between local peaks of order 5 to 10 micrometers. These are the result of the system's attempt to minimize energy -- when the equilibrium mean lattice spacing of an alloy being deposited is greater than the available mean lattice spacing of the exposed alloy surface, the system uses its degree of freedom in the $z$-direction to increase the mean spacing between deposited atoms. This yields coherent surface undulations in the [110] and [1-10] directions on the surface. For films deposited at sufficiently high temperature, sufficiently shallow alloy gradient, and at sufficiently low deposition rate, these undulations extend for thousands of micrometers. On films deposited under less optimal conditions, these undulations can be quite short, even 10 micrometers or less, at

which point their orientation becomes difficult to determine. Another indicator of poor quality is observed in films deposited with an excessive temperature schedule -- round pits appear in the surface. These are suspected to be due to gas-phase nucleation yielding particulate contamination of the surface and a resulting disruption of "uniform" epitaxial deposition.

Since the source and drain of the FETs are *n*-type, *p*-type doping for the body is needed. The substrate is thus boron doped, and diborane is flowed with the germane during formation of the germanium cap to yield a boron concentration of approximately $10^{17}/cm^3$ there. To effect good contact between the substrate and the FET bodies, it is also desirable to dope the graded-alloy region. Extensive work was done to achieve this. However, it was found that the use of diborane during the graded-layer formation reduced the deposition temperature at which surface pits, ... were formed. Additionally, chamber wall deposition

rate-limiting step in silane CVD, and since hydrogen bonds much more readily with silicon than with germanium, this process is effectively self-limiting -- silicon deposits on the exposed germanium surface but, once the surface is all silicon, hydrogen bonds with the surface and the growth is virtually blocked. Oxide deposition immediately follows this process.

The gate electrode is also formed in the epitaxial reactor. *In-situ* boron-doped $Ge_{0.30}Si_{0.70}$ is readily deposited at 500 C with a resulting resistivity of 1 mohm-cm. No further activation anneal is required. Deposition is initiated with a silicon seed layer. This is made thick enough (at least several extrinsic Debye lengths) to establish a well-defined workfunction at the electrode-insulator interface. Then, to avoid problems associated with band discontinuities, the germanium fraction is gradually graded up to 30%. After the bulk of the gate is thus deposited, the germanium fraction is continuously reduced back to zero and the growth is completed with a silicon capping layer, used to present a well-understood surface for later processing.

The remaining fabrication is standard silicon MOS -- implant $10^{15}/cm^2$ arsenic at 25 keV, activate the dopant at 500 C, deposit an LTO sub-metal dielectric, etch contact holes, and deposit and pattern titanium and aluminum sputtered metal. Finally, a 275 C forming gas anneal is done to improve the oxide-semiconductor interface and the conductivity of the metal-semiconductor contacts.

Initial testing of completed devices is expected to begin by the end of March 1996. Testing of strained-Ge devices is expected in April.

## Bibliography


[Daembkes] Daembkes et al; *IEEE TED*, **33**:663 1986.

[Fitzgerald] Fitzgerald et al; *APL* **59**:811-813 1991.

[Garone] Garone et al; *IEEE EDL* **12**(5):230-232 1991.

[Hymes] Hymes et al; *Journal of the Electrochemical Society*, **135**(4):961-965 1988.

[Ismail] Ismail et al; *IEEE EDL* **13**(5):229-231 1992.

[King] King et al; *IEEE EDL* **10**:52 1989.

[LeGoues91] LeGoues et al; *Phys Rev Letters*, **66**(22):2903-2906 1991.

[LeGoues92] LeGoues et al; *Journal of Applied Physics*, **71**:4230-4243 1992.

[Meyerson] Meyerson et al; *Applied Phys Letters*, **53**(25):2555-2557 1988.

[Nayak] Nayak et al; *IEEE EDL* **12**(4):154-156 1991.

[Pearsall86] Pearsall and Bean; *IEEE EDL* **7**:308 1986.

[Pearsall89] Pearsall; CRC Critical Reviews of in Solid State and Materials Sciences, **15**(6):551-600 1989.

[Schaffler] Schåffler et al; *Semiconductor Device Tech*, **7**:260-266 1992.

[Tsang] Tsang et al; *Appl Phys Let*, **62**(10):1146-1148 1993.

[Welser] Welser et al; *IEDM Tech Digest* 1000-1002 1992.

[Wang] Wang et al; *Materials Research Society Symp Proc*, **220**:403-408 1991.

[Xie] Xie et al; *Materials Research Society Symp Proc*, **220**:413-417 1991.

UNIT: 6

TITLE: Portable Video on Demand in Wireless Communication

PRINCIPAL INVESTIGATOR: T. H. Meng

GRADUATE STUDENT: K. Precoda

## I. Introduction

This research aims at providing low-power video compression for portable wireless video applications. We developed a power efficient video encoder architecture that uses pyramid vector quantization (PVQ) to compress video data. The decoded image quality using this encoder is better on average in terms of PSNR than JPEG.

In wireless communication, the available bandwidth generally changes with time. Our PVQ encoder, therefore, adjusts the frame rate according to the available bandwidth. If a large bandwidth is available, we increase the frame rate, improving the video quality at the receiver. If the bandwidth is limited, we decrease the frame rate, which results in degraded video quality. This ability to dynamically vary the compression rate allows the encoder to adaptively vary the amount of video data transmitted to achieve the best image quality for a given available bandwidth.

To handle variable frame rates while consuming the absolute minimal power, which is critical in portable systems, we propose to use circuits whose speed/power consumption can be adjusted by actual encoder throughput requirements. Our approach is to design a power supply controller that can adjust the DC voltage to control the desired performance. At high frame rates or when large bandwidth is available, the encoder would operate at high voltages, and, therefore, higher frequencies, allowing more image pixels to be processed per second. If smaller bandwidth is available, the supply voltage need not operate at a high voltage and is decreased appropriately to allow efficient operation at the required throughput. The encoder, therefore, consumes the absolute minimal power necessary to meet the frame rate of the encoder.

## II. Power-Supply Regulation

In order to provide a variable supply voltage as a function of the processing speed required, the voltage regulator must rapidly vary the supply voltage to meet the required throughput rate, while maintaining high power efficiency. We have designed a dc-dc switching regulator that

achieves efficiency in excess of 90% with a tracking speed of under 1 ms. The regulator supplies efficiently from a few milli-Watts to several hundred milli-Watts for all supply voltages of interest.

### A. Introduction to Switching Regulator

The switching regulator works by chopping the input battery voltage to generate a wave of pulses. These pulses pass through a second-order low-pass filter, which reduce the ac component to an acceptable ripple. The chopping is accomplished by active devices, which are integrated on a single chip to meet the size and weight requirements in portable applications. The inductor and capacitor, which form the low-pass filter, cannot be integrated to standard CMOS process, unfortunately, because of their large inductance and capacitance values. Consequently, off-chip inductors and capacitors are used.

### B. Low Power Techniques For Switching Regulators

The switching regulator can ideally achieve 100% efficiency. There are three main sources of dissipation which cause the conversion efficiency to be less than unity: conduction loss in the chopping transistors, switching loss due to parasitics, and gate drive loss.

To improve the conversion efficiency, we employ synchronous rectification and fixed pulse-width voltage modulation. A diode is typically placed between a ground and the input to the low-pass filter to drive the pulse to zero volt. For low-power applications, the voltage drop across the diode causes significant power loss compared to the power delivered. This conduction loss is minimized by replacing a diode with a gated NMOS, which reduces the conduction loss substantially. This use of NMOS is called *synchronous rectification.*

The output voltage is approximately equal to the input voltage multiplied by the duty factor. The duty cycle can be changed arbitrarily by varying the pulse-width or keeping the pulse-width constant and varying the operation frequency. Unlike most traditional switching regulators, we

performing appropriate feedback compensation techniques are well known. Since our encoder must operate at wide load conditions as well as operating voltages, the location of the poles and zeros move by substantial amounts. To maintain stability with a fast response time, the converter needs to track the large movements of poles and zeros and place the compensating poles appropriately. This complicates the controller, which increases power dissipation and lowers efficiency. A nonlinear feedback controller is, therefore, employed requiring only a few adders and comparators. This controller is shown to be stable for all operating regions of interest.

## III. Low-Power PVQ Encoder

## IV. Conclusion

The goal of this research was to study the energy-on-demand design methodology for implementing low-power video compression systems. The methodology introduced using our [illegible] processing applications, where the required throughput rates are time-variant. We are exploring other applications for this energy-on-demand design methodology.

## JSEP Supported Presentations

1. E. K. Tsern "A Low-Power Video-Rate Pyramid VQ Decoder," presentation at 1996 IEEE International Solid-State Circuits Conference, February 1996.
2. T. H.-Y. Meng, "A Low-Power Encoder Architecture for Pyramid Vector Quantization of 2-D Subband Coefficients," presentation at International Conference on Image Processing, October

5. B. M. Gordon, E. K. Tsern, and T. H.-Y. Meng, "Design of a Low-Power Video Decompression Chip Set for Portable Applications," invited submission to *Journal of VLSI Signal Processing*, October 1995.
6. W. Namgoong, M. Davenport, T. H.-Y. Meng, "A Low-Power Encoder Architecture for Pyramid Vector Quantization of 2-D Subband Coefficients," *Proceedings of 1995 IEEE Workshop on VLSI Signal Processing*, pp. 391-400, Osaka, Japan, October 1995.

UNIT: 7

TITLE: Adaptive DFE for GMSK in Indoor Radio Channels

PRINCIPAL INVESTIGATOR: J. M. Cioffi

GRADUATE STUDENTS: R. D. Wesel and K. Jacobsen

## I. Introduction

Point-to-multipoint transmission problems are finding increasing application in broadcast and data communication networks. Such problems were the main focus of the supported JSEP research. Two Ph.D. students are matriculating in 1996 in these areas, Richard Wesel and Krista Jacobsen. Both have significant results, as reported below, and several published or pending papers under this contract's support.

### Super-redundancy - R.D. Wesel

Rick Wesel's work focused on broadcast coding methods. In this area, a single source of digital information sends the same information to may remote users, with no feedback path. The transmission paths may vary from user to user and with time for a particular user. Such a situation is characteristic of terrestrial or satellite broadcast networks.

Rick found that to optimize a transmission system fully, the channel characteristic must be known to both the transmitter and the receiver. The consequent optimal action of the transmission system is then a function of this known channel characteristic. In the broadcast case, each user has a different channel characteristic and all are unknown to the transmitter. However, the maximum data rate that could be achieved by each of these users is roughly the same that should achieve at least the worst-case capacity on all the channels. Rick found this rate can be achieved without having to use different codes/designs for the different user paths.

Rick's work then progressed to a search for such a robust code, and several have been found as well as a general search procedure. These codes and the search procedure are described in Section II.

### Multipoint-to-point access protocol and analysis - K. Jacobsen

The main focus of Krista Jacobsen's research has been the mechanisms for upstream access in a point-to-multipoint transmission architecture. The specific architecture studied was tree-structured coaxial networks, but the results also apply to wireless and local-area networks.

This work has produced a number of protocols and contention resolution methods for multicarrier transmission with such networks. In particular, a combination of time and frequency division access are combined at the physical transmission layer to improve throughput versus latency trade-offs in such networks, as described in Section III.

A method for network synchronization and coordination was postulated for a multicarrier transmission system and reservation-based access protocols were investigated. Significant improvements in throughput and efficiency were obtained with respect to time-only multiplexing.

Both sets of work have resulted in a reasonable level of publication as reported in Sections II. and III.

## II. Trellis Codes for Correlated Fading - Rick Wesel

### The Problem

Consider transmission over one or more channels subject to fading in time or frequency such that the fading can be estimated at the receiver but is unknown to the transmitter. An

codes that are ideal for use in broadcast transmissions where a single transmission must work for a variety of different channels.

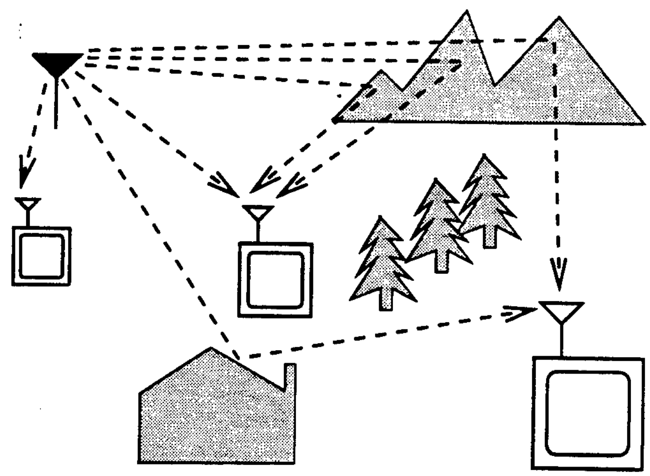


Figure 1: Digital Video Broadcast.

These new codes have already generated significant interest in industry. Telia Research, the Swedish telecommunications company, is exploring how these codes can be used to provide reliable wireless data links between a base station and a mobile user. Here again the transmitter cannot specialize the transmission to the particular fading. Unlike the broadcast situation, there is only one fading pattern. However, the transmitter does not know what that fading pattern is. Thus a robust code is required.

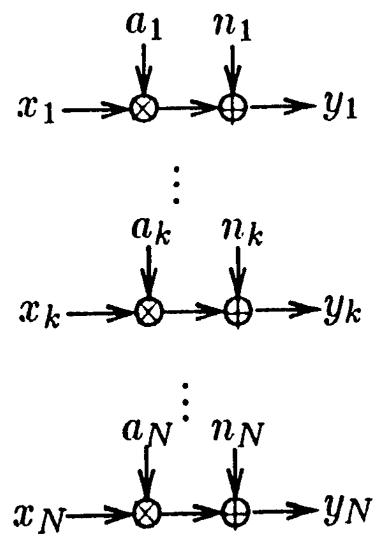



Figure 2: Overall subchannels with different SNRs.

**Super-Redundancy**

The first requirement, that the number of coded bits transmitted per symbol be large, implies that good fading--channel trellis codes will have a large amount of redundancy. This concept of super-redundancy can be contrasted with the additive white Gaussian noise channel, where it was shown that only one bit of redundancy is required [Ungerboeck]. In the fading environment, the subchannel capacities can vary by a large amount. To efficiently use the channel as a whole, each individual subchannel must be used efficiently. This requires that the number of coded bits be large enough that the high capacity subchannels can be fully utilized.

**Code Distance Distribution and Correlation in Fading**

[illegible] for fading channels should distribute distance

previous techniques, the permuted correlation in the interleaved fading channel is a primary consideration the code design procedure.

To utilize this correlation information in a straightforward way, periodic interleaving is used. The interleaving period is chosen small enough that symbols within one period are essentially uncorrelated. Symbols separated by multiples of the interleaving period are extremely correlated. Thus symbol--error distances on symbols separated by multiples of the interleaving period provide exactly one "diversity branch".

The code design search procedure finds the trellis code that spreads code distance as evenly as possible on as many of these diversity branches as possible. The number of diversity branches in such a scheme is upper bounded by the period of the interleaver. However, if this period is chosen correctly, that is also the limit of the diversity present in the fading environment. Detailed discussions of the code design procedure can be found in the publications listed at the end of this section.

### Performance of the New Codes

To see how well the new codes can perform we consider the example of multicarrier broadcast and consider the four different frequency responses shown in Fig. 3. A multicarrier system with 512 subcarriers in assumed and the desired information rate will be fixed at 1 bit per symbol. Our code design procedure produces a rate 1/4 convolutional code which is used to select points form a 16 QAM constellation. This code is compared with a standard code for multicarrier broadcast of 1 bit per symbol -- a rate 1/2 code used to select points from a 4 PSK constellation. Both codes have 64 states and thus require Viterbi decoders with the same complexity.

Figure 4 shows that the newly designed code provides consistent performance on all four of these channels. At a bit error rate of $10^6$ the new code has all four performance curves within a band of 0.75 dB. The standard code performs 1 dB better on the Flat Channel (Channel 1). However, it's performance becomes unacceptable as the frequency selectivity become more pronounced. On the Step Channel (Channel 4), which is a step in the frequency response, the standard code has bit error rates close to 1/2 for the entire range of the plot.

### Conclusion

The new codes produced by this research provide reliable performance over a wide variety of time/frequency fading patterns. This type of consistent reliability is unmatched by previous

techniques, and the new codes will find applications in numerous data communication applications including digital video broadcasting and wireless data networks.

## III. Design and Analysis of Multipoint-to-point Discrete-Multitone-based Networks - Krista S. Jacobsen

### The Problem

As the deployment of hybrid fiber-coax (HFC) networks by both cable television and telephone companies continues, efficient, cost-effective techniques to transmit digital multimedia signals both to and from the home must be developed. Transmission channels in the downstream direction (from the central site to the customer premise) are generally high-quality, and use of a single-carrier modulation in broadcast mode is probably sufficient for downstream transmission. However, the upstream bandwidth of HFC networks is often plagued by numerous transmission impairments, including passband ripple, spectral nulls, and radio-frequency ingress. Hence, a robust upstream modulation technique is required to ensure that effective communications can occur in the presence of these impairments. Furthermore, because HFC networks are generally configured in tree-and-branch topologies, as shown in Fig. 1, the return channel (upstream bandwidth) is shared among many users, potentially thousands. Consequently, use of the available upstream bandwidth must be coordinated somehow to ensure the channel is used efficiently.

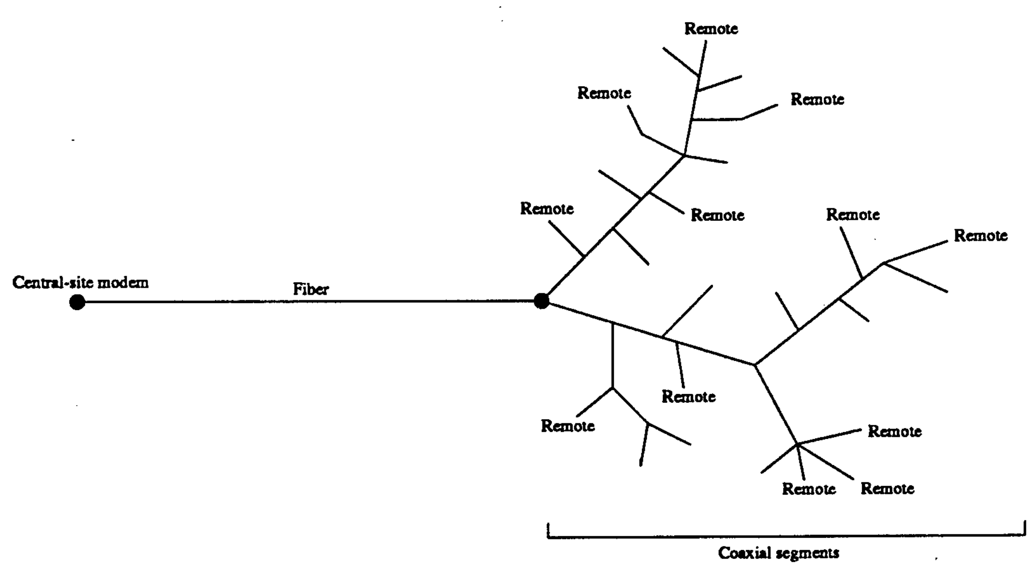



Figure 1: HFC network configuration.

Discrete-multitone (DMT), a type of multicarrier modulation, has been shown in previous JSEP-sponsored papers (i.e., [Jacobsen and Cioffi-a], [Jacobsen and Cioffi-b]) to offer significant advantages for upstream transmission in HFC channels, particularly because DMT can optimize the

HFC return channel. However, to exploit these advantages and achieve improvements in the

misalignment. The remote unit then implements the requested sample delay and transmits a signal requesting verification that it is synchronized. If the remote unit transmission is indeed synchronous, the central unit controller sends a signal to that unit in the downstream channel to indicate that no further shifting is required, and that the remote unit may now communicate with the central-site modem incorporating the appropriate delay. Otherwise, the synchronization procedure is repeated until the central-site controller determines the remote unit is synchronized. After the initial symbol delay has been determined, unless a remote unit is moved or its connection to the network is terminated, it should not have to be resynchronized. Failing to synchronize the remote units to within a certain tolerance can result in interchannel interference, which can decrease the achievable bit rates on the affected subchannels.

After receiving and incorporating the required sample delay from the central-site modem, an installing remote unit transmits a wide-band signal during a specified number of upcoming silent periods to train the central unit receiver. Because the newly installed remote unit is now synchronized with respect to the other remote units, it can transmit using all of the symbols during the next several silent periods for channel analysis. All other remote units remain quiet while the remote unit transmits a training signal on the permissible subset of the subchannels allocated to it, and the central unit controller records the bit capacity and magnitude and phase of each subchannel from that remote unit. The bit capacities are used to determine subchannel assignments when the remote later requests either a constant data rate or a packet transmission. Because the controller allocates the subchannels to the various remote units every symbol period, it can apply the appropriate subchannel magnitude/phase inverse to each subchannel to demodulate the received signal. Hence, if the remotes are all properly synchronized, the signal arriving at the central unit receiver, which is actually an aggregate of transmissions from a number of different remote units, can be demodulated as though it were from a single remote modem, using the appropriate mixture of subchannel magnitude/phase inverses.

After a remote has been installed, it is periodically retrained during another silent interval reserved specifically for this purpose. As during the installation silent period, all remote units that are not training remain quiet to allow the central unit controller to update its settings for the training remote. Depending on the frequency of these silent intervals, the number of remotes on a particular network, and other system parameters, each remote could be retrained as often as many times per second or as infrequently as every few seconds.

## Design and Analysis of the Reservation-Based Multicarrier (RBM) Protocol

After the remote units have been installed, synchronized, and trained, they are capable of transmitting without interfering with other remote units as long as they obey a channel access protocol. One alternative for controlling transmissions from remote units so that data is always transmitted collision-free is a reservation-based protocol. Under a generalized reservation-based protocol, to obtain permission to transmit data a remote unit must first transmit a reservation request. When a reservation has been granted, then the corresponding data message is guaranteed to be received intact (channel noise notwithstanding) by the central-site modem. If reservation requests are transmitted using the same bandwidth as data transmissions, then coordination of reservation requests is necessary to ensure they do not interfere with data transmissions.

The *Reservation-Based Multicarrier (RBM* ) [Jacobsen and Cioffi-d] protocol has been developed for multicarrier-based multipoint-to-point networks (such as HFC) in which data transmissions scheduled by a central controller are desirable because remote units are unable to detect whether or not the upstream channel is in use, and data transmissions and reservation requests occupy the same bandwidth. Under the RBM protocol, each multicarrier symbol is

holds. Regardless of what method is used to divide the subchannels into frequency-slots, the partitioning must be observed by all remote units.

After transmitting its RFB on one of the $K$ subchannel sets, a remote unit waits a specified period of time, determined by the round-trip propagation delay of the channel and the central unit processing time, to ascertain whether or not its RFB arrived successfully at the receiver. If the waiting remote does not receive a grant message from the central controller within a certain period of time, which indicates that its RFB collided with another unit's RFB or was unintelligible to the receiver for some other reason, it reschedules the RFB for a later time according to a delay distribution. If the remote does receive a grant message before timing out, it begins to transmit its message using all subchannels during the symbol period corresponding to the index sent by the central controller. Figure 2 illustrates the protocol timing, channel status signal, and upstream channel activity when a successful RFB occurs and the minimum delay is incurred.

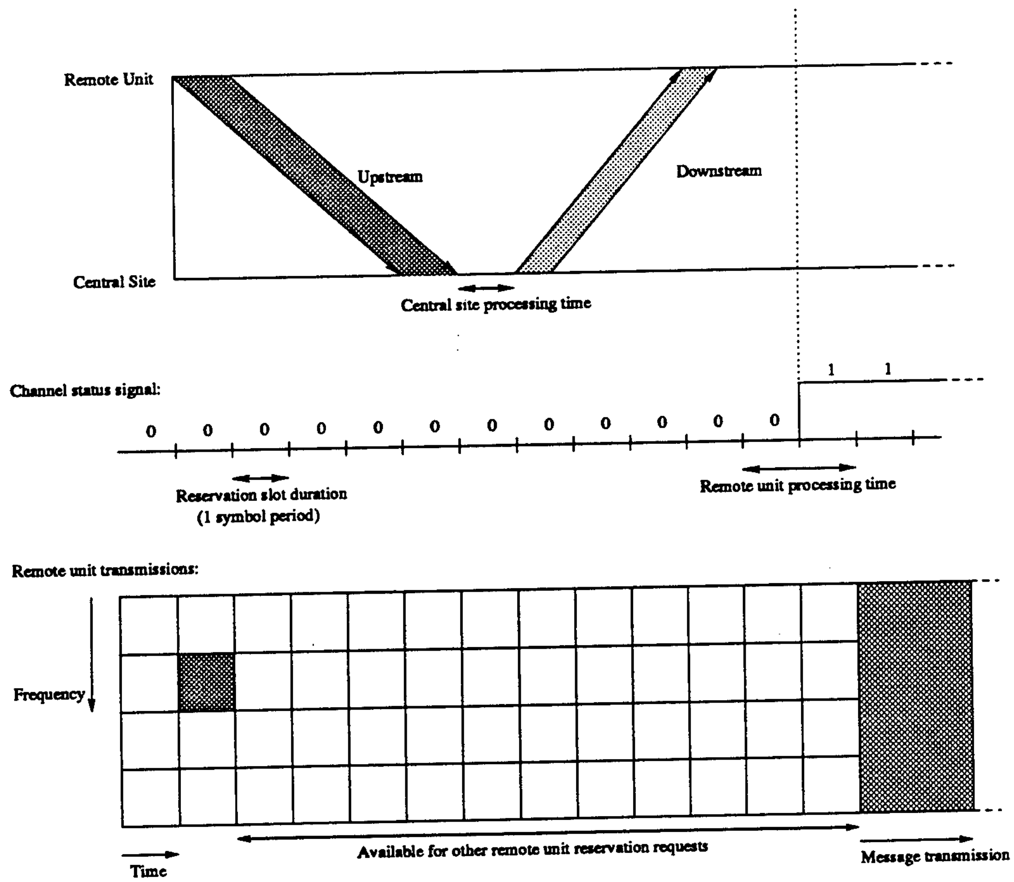
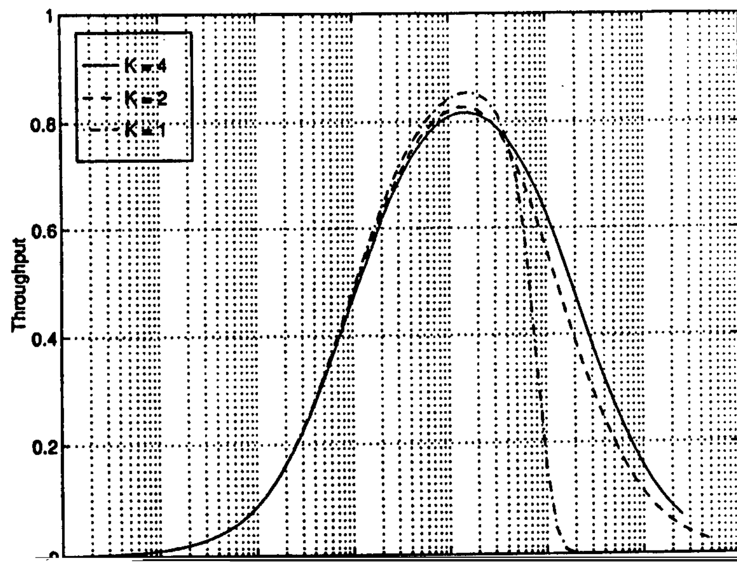



Figure 2: Illustration of protocol with K = 4.

To simplify protocol management, all multicarrier remote units are constrained to transmit using the same bit tables. In other words, for all remote units, the number of bits $b_i$ on the $i$ th subchannel is the same. Note that the number of bits supported by subchannel $i$ need not equal the number supported by subchannel $j$, as long as $b_i$ and $b_j$ are the same across all remote units on the network. Under the constraint of equivalent bit tables, the central unit receiver applies the same decoding procedure to every received symbol. Therefore, the receiver does not need to know in advance which of the remotes is transmitting an RFB or, for that matter, a message. Furthermore,



bandwidth was divided into 32 ($K = 4$), 16 ($K = 2$), and 8 ($K = 1$, slotted single-carrier) 2-bit subchannels. Hence, the time required to transmit each message is the same for each scenario, and the achievable throughputs for the various values of $K$ may be compared without modification.

The figure illustrates that the throughput achieved by the RBM protocol is a function of the

[Jacobsen and Cioffi-d] K. S. Jacobsen and J. M. Cioffi, "Achievable Throughput in Multicarrier-based Multipoint-to-point Networks Using a Reservation-based Channel Access Protocol," submitted to *Globecom '96*.

### JSEP Supported Publications


1. R. D. Wesel and J. M. Cioffi, "Fundamentals of Coding for Broadcast OFDM," In *Proceedings of the 29th Asilomar Conference on Signals Systems & Computers*, November 1995.
2. R. D. Wesel and J. M. Cioffi, "A Transmission System Using Codes Designed for Transmission with Periodic Interleaving," U. S. Patent Pending.
3. R. D. Wesel and J. M. Cioffi, "Trellis Codes for Channels with Correlated Fading," in Preparation for Submission to *IEEE Transactions on Communications*.
4. K. S. Jacobsen and J. M. Cioffi, "An Efficient Digital Modulation Scheme for Multimedia Transmission on the Cable Television Network," in *Technical Papers, 43rd Annual National Cable Television Association (NCTA) Convention and Exposition,* New Orleans, LA , May 1994.
5. K. S. Jacobsen and J. M. Cioffi, "High-performance Multimedia Transmissions on the Cable Television Network," in *Proceedings 1994 International Conference on*

UNIT: 8

TITLE: Robust Estimation Methods for Adaptive Filtering

PRINCIPAL INVESTIGATOR: T. Kailath

GRADUATE STUDENTS: Y. C. Pati and B. Hassibi

# 1 Introduction

Our earlier JSEP-supported work was concerned with the use of spatial and temporal (signal) structure in smart antennas for mobile radio networks. The work done there gradually led us to consider, and to study, the *robustness* of the underlying algorithms with respect to model uncertainties and lack of statistical information. In particular, of interest were adaptive filtering algorithms which are widely used in communications (as well as in many other areas) for the identification and equalization of channels.

Classical methods for such problems require a priori knowledge of the statistical properties of the signals. In many applications, however, one is faced with model uncertainties and lack of statistical information. Therefore the aforementioned methods are not directly applicable. Moreover, it is not even clear what the behaviour of such estimation schemes might be if the assumptions on the statistics and distributions are not exactly met.

Adaptive filtering techniques are currently widely used to cope with such model uncertainties and lack of a priori knowledge. The methods currently used fall into the two general classes of least-squares-based algorithms (such as recursive-least-squares or RLS) and gradient-based algorithms (such as least-mean-squares or LMS). While the former class is derived from an explicit cost function, it is suspect whether their robustness properties are always desirable. On the other hand, the former methods are rather ad-hoc and do not follow from a rigorous framework. However, the gradient algorithms are by far the ones most used in applications. Our work now provides some analytic explanation of this fact.

In the last decade such problems have received great attention in control theory, where a so-called $H^\infty$ approach has been extensively studied. It turns out, in particular, that the LMS algorithm is $H^\infty$-optimal, thus establishing the observed robustness of this very widely used algorithm. We have also obtained some results on the robustness of least-squares-based adaptive filters. This framework is currently being used to explore new adaptive filtering algorithms for nonstationary scenarios.

# 2 Adaptive Filtering

The standard model assumed in adaptive filtering is the following:

$$d_i = h_i^T w + v_i, \qquad i \geq 0 \tag{1}$$

where $\{d_i\}$ is an observed output sequence (often called the reference signal), $\{h_i\}$ is a known input vector sequence, $w$ is an unknown weight vector that we intend to estimate, and $\{v_i\}$ is an unknown disturbance, which may also include modeling errors. We shall make no assumptions on the statistics or distributions of the $\{v_i\}$.

We denote the estimate of the weight vector using all the information available up to time $i$ by

$$w_i = \mathcal{K}(d_0, d_1, \ldots, d_i; h_0, h_1, \ldots h_i).$$

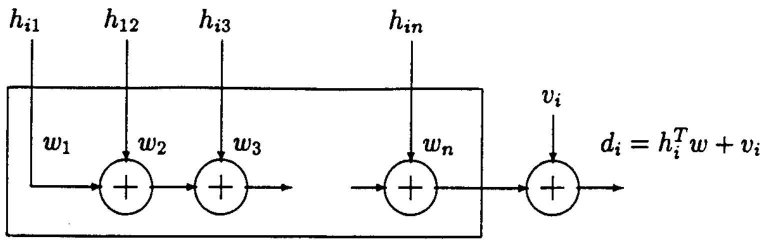



Figure 1: The model for adaptive filtering.

## 2.1 Least-Squares-Based Methods

There are a variety of choices for $w_i$, but the most widely used estimate $w_i$, is one that satisfies the following least-squares (or $H^2$ criterion):

$$\min_{w} \left[ \mu^{-1}|w - w_{-1}|^2 + \sum_{j=0}^{i} |d_j - h_j^T w|^2 \right], \tag{2}$$

where $w_{-1}$ is the initial estimate of $w$, and $\mu > 0$ represents the relative weight that we give to our initial estimate compared to the "sum of squared-error" term $\sum_{j=0}^{i} |d_j - h_j^T w|^2$. In the so-called *pure least-squares* problems one takes $\mu = \infty$, so that the first term in the cost function of (2) does not appear.

The *exact* solution to the above criterion is the RLS (Recursive Least Squares) algorithm:

$$w_i = w_{i-1} + k_{p,i}(d_i - h_i^T w_{i-1}) \quad , \quad w_{-1} \tag{3}$$

with $k_{p,i} = \frac{P_i h_i}{1 + h_i^T P_i h_i}$ and $P_{i+1} = P_i - \frac{P_i h_i h_i^T P_i}{1 + h_i^T P_i h_i}$, $P_0 = \mu I$.

RLS has certain stochastic optimality properties: if we assume in model (1) that the $w - w_{-1}$ and $\{v_i\}$ are zero mean independent Gaussian random variables with variances $\mu I$ and 1 respectively, then the RLS algorithm yields the maximum likelihood estimate of $w_i$. In particular, it minimizes the expected *prediction error energy*:

$$E \parallel e_p \parallel_2^2 = E \sum_{j=0}^{i} |h_j^T w - h_j^T w_{j-1}|^2. \tag{4}$$

## 2.2 Gradient-Based Algorithms

In gradient-based algorithms instead of exactly solving the least-squares problem (2), the estimates of the weight vector are updated along the negative direction of the *instantaneous* gradient of the cost function appearing in (2). Two examples are the LMS (Least-Mean-Squares)

$$w_i = w_{i-1} + \mu h_i(d_i - h_i^T w_{i-1}) \quad , \quad w_{-1} \tag{5}$$

and normalized LMS

$$w_i = w_{i-1} + \frac{\mu}{1 + \mu h_i^T h_i} h_i(d_i - h_i^T w_{i-1}) \quad , \quad w_{-1} \tag{6}$$

algorithms. Note that in the case of LMS the gain vector $k_{p,i}$ in RLS (which had to be computed by propagating a Riccati equation) has been simply replaced by $\mu h_i$. Likewise if we compare normalized LMS with the RLS algorithm, we see that the difference is that instead of propagating the matrix $P_i$ via the Riccati recursion we have simply set $P_i = \mu I$, for all $i$. Therefore the LMS and normalized LMS algorithms were long considered to be *approximate* least-squares solutions and were thought to lack a rigorous basis.

## 2.3 The Question of Robustness

We noted that under suitable stochastic assumptions, $H^2$-optimal adaptive filters have certain desirable optimality properties. However, a question that begs itself is what the performance of such filters will be if the assumptions on the disturbances are violated, or if there are modelling errors in our model so that the disturbances must include the modelling errors? In other words

- *is it possible that* **small** *disturbances and modelling errors may lead to* **large** *estimation errors?*

Obviously, a nonrobust algorithm would be one for which the above is true, and a robust algorithm would be one for which small disturbances lead to small estimation errors.

The problem of robust estimation is thus an important one. As we shall see in the next section, the $H^\infty$ robust estimation formulation is an *attempt* at addressing this question. The idea is to come up with estimators that minimize (or in the

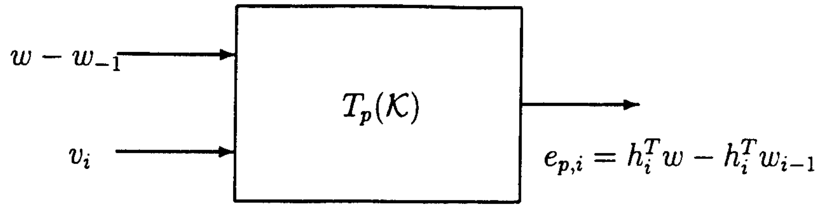



Figure 2: Transfer operator from the unknown disturbances $\{w - w_{-1}, v_i\}$ to the prediction errors $\{e_{p,i}\}$. Likewise for $T_f(\mathcal{K})$ and $T_s(\mathcal{K})$.

**Definition 1 (The $H^\infty$ Norm)** *The $H^\infty$ norm of a transfer operator $T$ is defined as*

$$\|T\|_\infty = \sup_{x \in h^2, x \neq 0} \frac{\|Tx\|_2}{\|x\|_2} \tag{8}$$

*where $h^2$ denotes the space of all square-summable causal sequences.*

We now propose to choose the estimator $\mathcal{K}$ so as to minimize the $H^\infty$ norms of $T_p(\mathcal{K})$, $T_f(\mathcal{K})$ and $T_s(\mathcal{K})$. To be more specific we have the following problem.

**Problem 1 ($H^\infty$ Adaptive Filtering Problem)** *Find estimators $w_i = \mathcal{K}_p(d_0, \ldots, d_i; h_0, \ldots, h_i)$, that minimize the maximum energy gain from disturbances to estimation errors for each of the aforementioned errors,* i.e. , *find estimation strategies $\mathcal{K}_p$, $\mathcal{K}_f$ and $\mathcal{K}_s$ such that*

$$\gamma_p^2 = \inf_{\mathcal{K}_p} \sup_{w,v \in h_2} \frac{\|e_p\|_2^2}{\mu^{-1}|w - w_{-1}|^2 + \|v\|_2^2} \tag{9}$$

$$\gamma_f^2 = \inf_{\mathcal{K}_f} \sup_{w,v \in h_2} \frac{\|e_f\|_2^2}{\mu^{-1}|w - w_{-1}|^2 + \|v\|_2^2} \tag{10}$$

*and*

$$\gamma_s^2 = \inf_{\mathcal{K}_s} \sup_{w,v \in h_2} \frac{\|e_s\|_2^2}{\mu^{-1}|w - w_{-1}|^2 + \|v\|_2^2} \tag{11}$$

*where $|w - w_{-1}|^2 = (w - w_{-1})^T(w - w_{-1})$ and $\mu > 0$ reflects a priori knowledge of how close $w_{-1}$ is to $w$.*

It turns out that nice solutions can be obtained for all three problems. The solutions to Prob. 1 are given below (see [Hassibia]), in which we have assumed that the input vectors $\{h_i\}$ are such that

$$\lim_{N \to \infty} \sum_{i=0}^{N} h_i^T h_i = \infty.$$

**Solution to (i):** If $\mu$ satisfies the bound

$$0 < \mu < \inf_i \frac{1}{h_i^T h_i} \tag{12}$$

then $\| T_p(\mathcal{K}) \|_\infty$ is minimized by the LMS algorithm with learning rate $\mu$,

$$w_i = w_{i-1} + \mu h_i (d_i - h_i^T w_{i-1}) \quad , \quad w_{-1}$$

and the minimum $H^\infty$ norm is given by

$$\gamma_p = 1.$$

*Remarks:*

(a) The fact that $\gamma_p = 1$ indicates that there is no amplification of the disturbances. Thus the prediction error energy will never exceed the disturbance energy.

(b) The above result is true only of the learning rate $\mu$ satisfies the bound (12). This is in accordance with the well-known fact that LMS behaves poorly of the learning rate is chosen too large.

**Solution to (ii):** $\| T_f(\mathcal{K}) \|_\infty$ is minimized by the normalized LMS algorithm

$$w_i = w_{i-1} + \frac{\mu}{1 + \mu h_i^T h_i} h_i(d_i - h_i^T w_{i-1}) \quad , \quad w_{-1}$$

and the minimum $H^\infty$ norm is given by

$$\gamma_f = 1.$$

*Remark:* Note once more that there is no amplification of the noise. Now, however, we have no restriction on $\mu$.

**Solution to (iii):** $\| T_s(\mathcal{K}) \|_\infty$ is minimized by the least-squares solution, and the minimum $H^\infty$ norm is

$$\gamma_s = 1.$$

*Remark:* Thus least-squares algorithms are $H^\infty$ optimal with respect to smoothing errors.

# 4 Robustness of Least-Squares Algorithms

Now that we have developed the $H^\infty$ optimality of the LMS and normalized LMS algorithms with respect to prediction and filtered errors, it is natural to ask what the performance of the RLS algorithm will be with respect to these error criteria.

In order to answer the above question we need to compute the $H^\infty$ norm of the RLS algorithm. Finding this $H^\infty$ norm essentially amounts to finding the maximum singular value of a linear time-varying operator. Upper bounds on the $H^\infty$ norm can be found by checking for the positivity of the solution of a certain time-varying discrete-time Riccati recursion. Although both approaches can be used in principle, they require knowledge of *all* the input data vectors $\{h_i\}$.

Since in adaptive filtering problems we are given, and are forced to process, the data in real time, we cannot store all the data and use the aforementioned methods to compute bounds for the $H^\infty$ norm. Therefore the main effort in the results given below is to obtain bounds on $H^\infty$ norm that use simple a priori knowledge of the $\{h_i\}$ and not their explicit values [Hassibib].

(i) For RLS, we can show

$$(\sqrt{R} - 1)^2 \leq \sup_{w,v \in h_2} \frac{\|e_p\|_2^2}{\mu^{-1}|w - w_{-1}|^2 + \|v\|_2^2} \leq (\sqrt{R} + 1)^2$$

or to give a "looser" bound

$$(\sqrt{1 + \mu \underline{h}^2} - 1)^2 \leq \sup_{w,v \in h_2} \frac{\|e_p\|_2^2}{\mu^{-1}|w - w_{-1}|^2 + \|v\|_2^2} \leq (\sqrt{1 + \mu \bar{h}^2} - 1)^2,$$

where

$$R = \max_i 1 + h_i^T P_i h_i \quad , \quad \bar{h}^2 = \max_i |h_i|^2 \quad , \quad \underline{h}^2 = \min_i |h_i|^2 .$$

**Remark:** Note that for large $\mu$, the $H^\infty$ norm grows as $\mu$. This shows that the pure least-squares problem (with $\mu = \infty$) is highly non-robust with respect to prediction errors.

(ii) For filtered errors we have

$$\sup_{w,v \in h_2} \frac{\|e_f\|_2^2}{\mu^{-1}|w - w_{-1}|^2 + \|v\|_2^2} \leq (\sqrt{1/r} + 1)^2 \leq 4,$$

where

$$r = \min_i 1 + h_i^T P_i h_i \geq 1.$$

**Remarks:**

(a) Note that, as with normalized LMS, the $H^\infty$ norm does not depend on $\mu$.

(b) The above result for filtered errors is an intermediate stage between the smoothing error case (where the $H^\infty$ and $H^2$ optimal filters coincide) and the prediction error case (where the performance of LMS and RLS can be drastically different.)

# 5 Future Work

The $H^\infty$ approach to adaptive filtering described in the previous section suggests several directions for future research. We mention a few here.

## 5.1 Time-Varying Problems

So far we have assumed that the weight vector, $w$, is constant in time. In many applications one needs to assume a time-varying, $w$, and must therefore devise algorithms that can track the time-variations of the weight vector.

In such cases, one approach is to use windowing. Two common windowing schemes are the following.

(i) **Exponential Window:** The exponential window gives (exponentially) larger weight to the more recent data. In particular, the prediction error and disturbance energies are computed as:

$$\sum_{j=0}^{i} \lambda^{-j} |e_j|^2 \quad \text{and} \quad \sum_{j=0}^{i} \lambda^{-j} |v_j|^2, \tag{13}$$

where $0 < \lambda < 1$ is the so-called forgetting factor that is chosen based upon a priori knowledge of how fast the weight vector varies with time.

(ii) **Finite-Memory Window:** In this case one only considers the last $L$ data points so that the prediction error and disturbance energies are computed as

$$\sum_{j=i-L+1}^{i} |e_j|^2 \quad \text{and} \quad \sum_{j=i-L+1}^{i} |v_j|^2, \quad \text{respectively.} \tag{14}$$

$L$ is often referred to as the window length.

It is therefore useful to consider the $H^\infty$ filters that result from such "windowed" definitions of energy. The filters that are obtained in this fashion will have good tracking properties and, at the same time, be robust.

## 5.2 Mixed $H^2/H^\infty$ Estimation

Fig. 5.2 shows the (squared) singular values of $\mathcal{T}_{p,rls}$ and $\mathcal{T}_{p,lms}$ (the transfer operators from disturbances to estimation errors for RLS and LMS) for $N = 50$ (where $N$ is the number of observed data points) and $\mu = .9$, for a simple one-dimensional adaptive filtering problem. As can be seen the maximum singular value for $\mathcal{T}_{p,lms}$ is one, whereas for $\mathcal{T}_{p,rls}$ it is much larger. On the other hand, the RLS algorithm minimizes the Frobenius norm (the sum of the squared singular values) of the transfer operator $\mathcal{T}_K$ which can be visualized as the area under the curve of the (squared) singular values. Thus if we choose disturbances uniformly from the space $\mathcal{C}^{50}$, the RLS algorithm will have better average performance than LMS, although its worst-case performance is significantly worst.

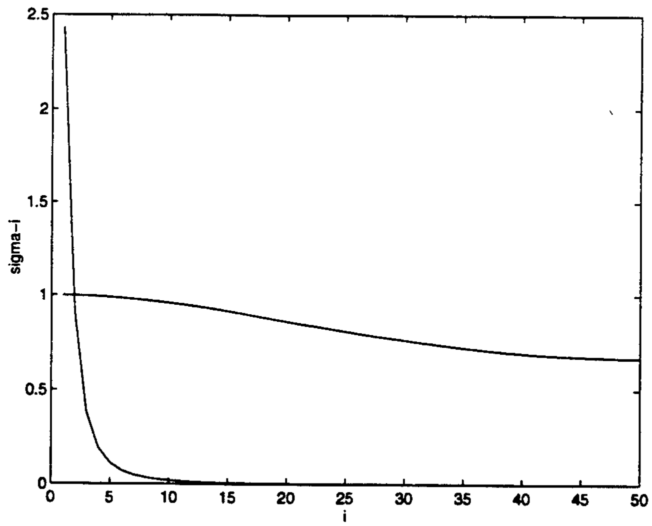



Figure 3: Singular values for $\mathcal{T}_{p,rls}$ and $\mathcal{T}_{p,lms}$ for $N = 50$ and $\mu = .9$.

Note, moreover, that although the LMS algorithm does not allow any amplification of the disturbances, it does not provide significant suppression of the disturbances, either. (The smallest squared singular value for $\mathcal{T}_{p,lms}$ which represents the minimum energy gain is roughly 0.65.) Since the $H^\infty$ optimal filters are not unique (LMS is only the central solution), it is very interesting to study the possibility of choosing other $H^\infty$ optimal filters to further reduce the Frobenius norm of $\mathcal{T}_K$. This will result in algorithms that have the best possible average behaviour while at the same time having the best possible worst-case performance. This framework is called the mixed $H^2/H^\infty$ estimation framework and is an area that we intend to pursue.

# References

[Hassibia] B. Hassibi, A.H. Sayed and T. Kailath. $H^\infty$ Optimality of the LMS Algorithm. *IEEE Trans. on Signal Processing*, vol 44, pp. 267-281, February 1996.

[Hassibib] B. Hassibi and T. Kailath. $H^\infty$ bounds for the recursive-least-squares algorithm. in *Proceedings of the 33rd IEEE Conference on Decision and Control*, pp. 3927-3929, Orlando. FL, Dec 1994.

UNIT: 9

TITLE: Efficient Data Compression

PRINCIPAL INVESTIGATOR: T. Cover

GRADUATE STUDENTS: E. Erkip, P. Fahn, G. Iyengar,

# 3 Detailed Research Descriptions

## 3.1 Image Compression

Our experiment to compare the image compression abilities of humans and computers is in its final stages. Our goal is to estimate the minimal rate, in bits per pixel, at which an image can be compressed without incurring significant perceptible distortion. First, one experimental subject simplies a given image without significantly distorting it, and then another subject predicts the simplified image, pixel by pixel, as accurately as possible. The accuracy of the second subject's predictions can be quantified to yield an estimate of the entropy of the simplified image. Not only will our results be useful as a benchmark to researchers in the field, but the experimental framework itself may lead to a new algorithm for data compression. A paper detailing the results of the experiment is currently under preparation.

## 3.2 Voice Channel

The thrust of this research is to develop a characterization of the capacity and optimal coding vocabularies of voice channels, which are mathematical models intended to capture properties of human speech generation. This research area will provide guidance on data compression for a voice channel or other channels with similar characteristics.

Consider a communication system with a channel characterized by a linear filter $g$ in an additive Gaussian noise environment, i.e., $y(t) = u(t) * g(t) + z(t)$, where $z(t)$ is white noise and $u(t)$ is the channel input. Instead of fixing the filter $g(t)$, which is the traditional approach, we fix the input signal $u(t)$ and attempt to choose a distribution on the space of linear, passive, causal filters $g(t)$ that maximizes the mutual information between the output and the filter. This model and its discrete-time analog are, we propose, an approximate model for the voice generation process

## 3.3 Feedback in Communication

It was recently shown by [Pombra and Cover] that the maximum achievable throughput (sum of rates of all users) of a Gaussian multiple access channel with feedback is at most twice that achievable without feedback. We prove [Ordentlich] a somewhat stronger result which establishes the factor of two bound not only for the total throughput but for the entire capacity region as well. Specifically, we show that the capacity region of a Gaussian multiple access channel with feedback is contained within twice the capacity region without feedback.

We have recently extended the factor of two bound on the capacity region of Gaussian multiple access channels to channels with inter-symbol interference (ISI). For single user Gaussian channels there is no information theoretic complication introduced by the addition of a causal linear filter at the transmitter. If the filter is invertible, the channel can be transformed into an ISI-free channel with an appropriately modified noise spectrum. For the multiple access channel, if the ISI filters are not identical for all transmitters, as is the case in practice, no such transformation is possible. This new result demonstrates that in wireless communications networks, once steady state has been reached via power control and channel learning, the maximum additional gain in capacity region afforded by receiver-to-transmitter feedback is limited to a factor of two, no matter how cleverly the feedback is used.

## 3.4 Robustness of Communication

Lapidoth, in a series of papers [Lapidoth 1], [Lapidoth 2], [Lapidoth 3], has considered the robustness of signaling in the presence of noise in an unknown environment. It is well known that Gaussian signals and matched filter decoding is optimal for signaling with a power

compression. Furthermore, one can compress partly entangled pairs of quantum particles into a small number of completely entangled pairs — the so-called Bell states — which can then be used for efficient communication of quantum data.

## References


[Castelli and Cover] V. Castelli and T. Cover. On the Exponential Value of Labeled Samples. *Pattern Recognition Letters*, 16:105-111, January 1995.

[Cover and King] T. Cover and R. King. A Convergent Gambling Estimate of the Entropy of English. *IEEE Trans. on Information Theory*, IT-24(4):413–421, July 1978.

# 4 Publications Supported by JSEP

## 4.1 Ph.D. Theses Supported by JSEP

A. Lapidoth, "Mismatched Decoding of the Multiple-Access Channel and Some Related Issues in Lossy Source Compression," August 1995.

## 4.2 Published Papers Supported by JSEP

1. S. Pombra and T. Cover. Non-White Gaussian Multiple Access Channels with Feedback, *IEEE Transactions on Information Theory*, 40(3):885-892, May 1994.
2. Z. Zhang and T. Cover. On the Maximum Entropy of the Sum of Two Dependent Random Variables. *IEEE Transactions on Information Theory*, 40(4):1244-1246, July 1994.
3. V. Castelli and T. Cover. On the Exponential Value of Labeled Samples. *Pattern Recognition Letters*, 16:105-111, January 1995.

## 4.3 Papers Submitted for Publication

1. T. Cover and E. Ordentlich. Universal Portfolios with Side Information. To appear in